

Aus der Klinik für Pferde  
der Veterinärmedizinischen Fakultät der Universität Leipzig

**Metaanalyse klinischer Studien 1983-2016 zur langfristigen Gebrauchsfähigkeit  
von Sportpferden nach Behandlung von natürlich entstandenen Erkrankungen der  
oberflächlichen und der tiefen Beugesehne und des Fesselträgers  
entweder allein mit kontrollierter Bewegung oder kombiniert  
mit einem potenziell regenerativen Therapeutikum**

Inaugural-Dissertation  
zur Erlangung des Grades eines  
Doctor medicinae veterinariae (Dr. med. vet.)  
durch die Veterinärmedizinische Fakultät  
der Universität Leipzig

eingereicht von  
Sarah Eva Marie Doll  
aus Mannheim

Leipzig 2019

Mit Genehmigung der Veterinärmedizinischen Fakultät der Universität Leipzig

Dekan: Prof. Dr. Walter Brehm

Betreuer: Prof. Dr. Walter Brehm

Gutachter: Prof. Dr. Walter Brehm, Klinik für Pferde, Veterinärmedizinische Fakultät, Leipzig  
PD Dr. Florian Geburek, Klinik für Pferde, Justus-Liebig-Universität, Gießen

Tag der Verteidigung: 05.03.2019

**Meinen Eltern**

# INHALT

1	EINLEITUNG	.	.	.	.	1
2	LITERATURÜBERSICHT	.	.	.	.	2
2.1	<b>Definition von prT und therapeutisches Konzept</b>	.	.	.	.	2
2.2	<b>Merkmale und Einflussfaktoren des langfristigen Behandlungserfolgs mit prT</b>					3
2.2.1	Merkmale	.	.	.	.	3
2.2.2	Einflussfaktoren	.	.	.	.	4
2.3	<b>Die Metaanalyse als Modell zur Bewertung klinischer Studien</b>	.	.			6
2.3.1	Die Metaanalyse als Methode der evidenzbasierten Veterinärmedizin	.	.			6
2.3.2	Drei Typen der Metaanalyse	.	.	.	.	7
2.3.3	Die Metaanalyse in der equinen Regenerationsmedizin	.	.	.	.	8
3	FRAGESTELLUNG, ZIELE UND HYPOTHESEN	.	.	.	.	10
4	MATERIAL UND METHODEN	.	.	.	.	11
4.1	<b>Literatursuche</b>	.	.	.	.	11
4.1.1	Ein- und Ausschlusskriterien	.	.	.	.	11
4.1.2	Online Recherche	.	.	.	.	11
4.1.3	Manuelle Recherche	.	.	.	.	11
4.1.4	Suchstrategien in der Datenbank PubMed	.	.	.	.	11
4.1.5	Suchstrategien in der Datenbank CAB Abstracts	.	.	.	.	12
4.1.6	Suche nach Metaanalysen in der Pferdemedizin	.	.	.	.	13
4.2	<b>Arbeitsmethoden</b>	.	.	.	.	13
4.2.1	Arbeitsschritte in Teil I: Extraktion der Behandlungsgruppen - qualitative und quantitative Analyse	.	.	.	.	13
4.2.2	Arbeitsschritte in Teil II: Methodisch systematische Entwicklung des <i>GFR</i> -Modells	.	.	.	.	14
4.2.3	Statistische Methoden	.	.	.	.	14
4.2.3.1	Computerprogramme	.	.	.	.	14
4.2.3.2	Statistische Heterogenitätsanalyse	.	.	.	.	14
4.2.3.3	Fixed Effects–Modell	.	.	.	.	15
4.2.3.4	Random Effects–Modell	.	.	.	.	16
4.2.3.5	Quantitative Kriterien der Heterogenitätsanalyse	.	.	.	.	17
4.2.3.6	Power-Analyse der Effektdifferenz zwischen Behandlungsgruppen.	.	.	.	.	18
4.2.4	Reduzierung von Fehlerquellen und Bias-Risiken	.	.	.	.	19
4.2.4.1	Typische Fehler in Metaanalysen	.	.	.	.	19
4.2.4.2	Vier häufige Formen von Bias in den Studien	.	.	.	.	20
4.2.4.3	Randomisierte kontrollierte klinische Studien – RCTs	.	.	.	.	20
4.2.4.4	Nicht randomisierte kontrollierte bzw. nicht randomisierte nicht kontrollierte klinische Studien	.	.	.	.	20

<b>5</b>	<b>ERGEBNISSE</b>	22
	TEIL I – METAANALYSE DER STUDIEN 2011-2016	26
<b>5.1</b>	<b>Behandlungsgruppen (BehG) in den Studien 2011-2016</b>	26
5.1.1	Einflussfaktoren und interessierende Merkmale des Behandlungserfolgs	26
<b>5.2</b>	<b>Deskriptive Vergleiche zwischen den Studien im zeitlichen Verlauf 1983-2011 und 2011-2016</b>	27
5.2.1	Häufigkeitsverteilung der BehG, Studien, Behandlungsformen (PrTB, KtrB) und Patienten in 1983-2011 und 2011-2016	27
	Zwischenergebnis	
5.2.2	Variabilität der Patienten-Fallzahl in den BehG	28
	Zwischenergebnis	
5.2.3	Verteilung der Patienten-Fallzahlklassen in den BehG	29
	Zwischenergebnis	
<b>5.3</b>	<b>Vergleich der Gewebequellen und prT in den Studien 2001-2016</b>	29
5.3.1	19 verschiedene prT	29
5.3.2	Verteilung der 19 verschiedenen prT differenziert nach autologer, allogener und xenogener Gewebequelle	30
5.3.3	Zunehmende Diversifizierung der prT in den Studien im Verlauf 2001-2011 und 2011-2016	31
	Zwischenergebnis	
5.3.4	Verteilung der Studien differenziert nach autologen, allogenen, xenogenen prT	31
	Zwischenergebnis	
5.3.5	Verteilung der Studien differenziert nach Knochenmark, Fettgewebe, Amniongewebe und Blut als Gewebequelle für MSC, sowie nach Blutprodukten und UBM	32
	Zwischenergebnis	
5.3.6	Absolute Anwendungshäufigkeit der insgesamt 9 Gewebequellen in Studien 2001-2016	33
<b>5.4</b>	<b>Behandlungsgruppen mit Angabe der 'Gebrauchsfähigkeitsrate' GFR</b>	33
5.4.1	Arbeitsschritte der Analyse der 50 BehG mit Angabe der GFR	34
5.4.2	Heterogenitätsanalyse zur Behandlungsform PrTB	35
5.4.2.1	Kriterien für die Poolbarkeit von BehG	35
5.4.2.2	Heterogenitätsanalyse der 11 Behandlungsgruppen-Paare zu PrTB	37
5.4.2.2.1	Bei 4 der 11 Paare sind beide BehG paarintern poolbar.	38
5.4.2.2.2	Bei 7 der 11 Paare sind beide BehG paarintern nicht poolbar	41
	Ergebnis der Heterogenitätsanalyse für 11 Paare zu PrTB	41
5.4.3	Heterogenitätsanalyse zur Behandlungsform KtrB	42
5.4.3.1	Heterogenitätsanalyse der 4 BehG zu KtrB	43
	Ergebnis der Heterogenitätsanalyse für 4 BehG zu KtrB	46
5.4.3.1.1	Vergleich zwischen Rennpferden und Nicht-Rennpferden	46

5.4.4	Power-Analyse der Effektdifferenz $d_{GFR}$	48
5.4.4.1	Untersuchung der BehG mit Angabe der $GFR$ auf BehG, die sich möglicherweise für paarweise Vergleiche (pV) eignen	48
5.4.4.1.1	'24 alleinstehende' Behandlungsgruppen zu PrTB Zwischenergebnis	49
5.4.4.2	Systematische quantitative und qualitative Analyse der 31 BehG auf adäquate pV Zwischenergebnis	53
5.4.4.3	Ergebnisse der quantitativen Power-Analyse für adäquate pV	55
5.4.5	Zusammenfassung der Ergebnisse der Analyse der 50 BehG mit Angabe der $GFR$	57

Teil II – DAS EVALUATIONSMODELL $GFR$	58
---------------------------------------	----

<b>5.5</b>	<b>Erfolgsmerkmal, Effektgröße und Störfaktoren des Modells</b>	58
5.5.1	Definition des Erfolgsmerkmals	58
5.5.2	Definition der 'Gebrauchsfähigkeitsrate' $GFR$ aus $E$ und $R$	58
5.5.3	Normierung der quantitativen Effektgröße $GFR$	59
5.5.3.1	Definition der quantitativen Variablen $A, B, C$	59
5.5.3.2	Beschreibung der disjunkten Fallgruppen $A, B, C$	59
5.5.3.3	Definition von $E, R, GFR$ durch die Variablen $A, B, C$	61
5.5.4	Störfaktorfremie Zusammensetzung der Studienpopulation	61
<b>5.6</b>	<b>Qualität der diagnostische Parameter</b>	62
5.6.1	Ganganalyse, Entzündungsparameter, Palpationsparameter	62
5.6.2	B-Mode-Ultraschalldiagnostik; Messtechnik des UTC-imaging	63
5.6.3	Läsionsgrad nach RANTANEN et al. (2011)	63
5.6.3.1	Semiquantitative Parameter: Echogenität; FAS	64
5.6.3.2	Quantitative Parameter: CSA-Differenz, MIZ-CSA, T-lesion %	65
5.6.3.3	Verletzungskategorien: Läsionsgrade (LG)	65
<b>5.7</b>	<b>Kontrolle der Messbedingungen zur Bestimmung von <math>E</math> und <math>R</math></b>	
	<b>Standardisierung der Studiendauer</b>	67
5.7.1	Rehabilitationsdauer	67
5.7.2	Variabilität der 'individuellen Rehabilitationsdauer'	68
5.7.3	Vorläufiger Schätzwert für die 'projektierte Rehabilitationsdauer'	68
5.7.3.1	Analyse der Rehabilitationsdauer zu KtrB bei OBS-TBS vs. FT	68
5.7.3.2	Analyse der Rehabilitationsdauer zu PrTB bei OBS-TBS vs. FT	70
5.7.4	'Projektierte Rehabilitationsdauer' gemäß $GFR$ -Modell	72
	<b>Exkurs: Bewegungsprogramme nach GILLIS (1997) und SMITH (2016)</b>	73
5.7.5	Prüfzeit	75
5.7.6	Kriterien der Standardisierung der Prüfzeit	75
5.7.7	Analyse der Prüfzeit bei OBS-TBS und FT	75
5.7.8	Vergleich der Prüfzeit zwischen OBS-TBS vs. FT	77
5.7.9	Prüfzeit gemäß $GFR$ -Modell	78
5.7.10	Standardisierung der Studiendauer gemäß $GFR$ -Modell	78
5.7.11	Zusammenhang zwischen Zunahme der Prüfzeit und Abnahme der $GFR$	79
	<b>Exkurs: Rennleistung und Prüfzeit</b>	79

<b>5.8</b>	<b>Ausblick</b>	.	.	.	.	80
5.8.1	Das gemeinsame Effektgrößen-Paar der Form ( <i>GFR</i> / Rehabilitationsdauer)	.	.	.	.	80
5.8.2	Die Norm- <i>GFR</i> der Behandlungsform KtrB	.	.	.	.	81
5.8.3	Modellevaluierung und Weiterentwicklung	.	.	.	.	81
<b>6</b>	<b>DISKUSSION</b>	.	.	.	.	82
<b>6.1</b>	<b>Bisherige Bewertungspraxis</b>	.	.	.	.	82
6.1.1	Rezidivrate	.	.	.	.	82
	<i>Beispielfälle 1- 2</i>					
6.1.2	Rehabilitationsdauer, Prüfzeit und Erfolgsmerkmal	.	.	.	.	83
	a) Das Erfolgsmerkmal 'recovery rate' mit der festgelegten 'Rehabilitationsdauer ein Jahr'	.	.	.	.	83
	<i>Beispielfall 3</i>					
	b) Das Erfolgsmerkmal: 'Rückkehr zum ursprünglichen Leistungsniveau nach einem beschleunigten Rehabilitationsprogramm von 5 Monaten und anschließend Rezidiv frei für mindestens 1 Jahr'	.	.	.	.	83
	<i>Beispielfall 4</i>					
	c) Unscharfe Definition des Erfolgsmerkmals	.	.	.	.	84
	<i>Beispielfall 5</i>					
6.1.3	Lückenhafte Präsentation der Quelldaten	.	.	.	.	85
	<i>Beispielfälle 6-7</i>					
<b>6.2</b>	<b>Diversifizierung der prT</b>	.	.	.	.	86
<b>6.3</b>	<b>Hohe bzw. niedrige Evidenz der Ergebnisse</b>	.	.	.	.	86
6.3.1	Amniongewebe-MSC	.	.	.	.	86
6.3.2	KtrB	.	.	.	.	87
6.3.3	PRP	.	.	.	.	87
<b>7</b>	<b>ZUSAMMENFASSUNG</b>	.	.	.	.	88
<b>8</b>	<b>SUMMARY</b>	.	.	.	.	90
<b>9</b>	<b>LITERATURVERZEICHNIS</b>	.	.	.	.	93
	DANKSAGUNG	.	.	.	.	104

## Abkürzungsverzeichnis

<i>a</i>	Jahr
AAEP	American Association of Equine Practitioners
ACP	autologes konditioniertes Plasma
ACS	autologes konditioniertes Serum
Ad-MSC	adiposed derived MSC
AP	autologes Plasma
BehG	Behandlungsgruppe
BM-MSC	bone marrow derived MSC
CSA	cross-sectional area
<i>E</i>	Erfolgsrate
EBM	Evidenzbasierte Medizin
EBVM	Evidenzbasierte Veterinär Medizin
EM	Erfolgsmerkmal
ES	echo score
FAS	fiber alignment score
FE	Fixed-Effects Modell
FT	Fesselträger
GF	langfristige Gebrauchsfähigkeit
<i>GFR</i>	Gebrauchsfähigkeitsrate
IL	Interleukin
IQR	Inter-Quartils-Abstand
k. A.	Keine Angaben
KtrB	Behandlung mit kontrollierter Bewegung allein
LG	Läsionsgrad
<i>M</i>	Monat
MIZ	maximal injury zone
MSC	mesenchymale stromale Progenitorzellen
Mw	Mittelwert
OBS	oberflächliche Beugesehne
P	Prüfzeit
PC	Thrombozytenkonzentrat
PRS	Thrombozytenreiches Serum
PRP	Thrombozytenreiches Plasma
prT	potenziell regeneratives Therapeutikum
PrTB	Behandlung mit prT kombiniert mit kontrollierter Bewegung
pV	paarweiser Vergleich
Q <sub>i</sub>	Quartil
<i>R</i>	Rezidivrate
RCT	randomised controlled trial
RE	Random-Effects Modell



STD	Standardabweichung
TBS	tiefe Beugesehne
T-ES	total echo score
T-FAS	total fiber alignment score
UBM	porcine urinäre bladder matrix
USB	Unterstützungsband
95%-CI	95-Prozent-Konfidenzintervall

## Abbildungsverzeichnis

Abb. 4.1.5	Suchstring in der Datenbank CAB Abstracts	.	.	13
Abb. 5.4.3.1	Forest-Plot und Funnel-Plot zu KtrB bei 3 BehG (N-A =30 Sportpferde)	.	.	44
Abb. 5.4.3.1.1a	Forest-Plot und Funnel-Plot zu KtrB bei 3 BehG (N-A =16 Rennpferde)	.	.	46
Abb. 5.4.3.1.1b	Forest-Plot und Funnel-Plot zu KtrB bei 6 BehG (N-A = 44 Nicht-Rennpferde)	.	.	47

## Tabellenverzeichnis

Tab. 4.1.1	Einschlusskriterien, Ausschlusskriterien . . . . .	11
Tab. 5.1a	Vergleich von Studien (2011-2016) zur Behandlung natürlich entstandener Erkrankungen des Sehnenapparates der OBS, TBS, USB der OBS mit kontrollierter Bewegung allein oder kombiniert mit potenziell regenerativen Therapeutika . . . . .	22
Tab 5.1b	Vergleich von Studien (2011-2016) zur Behandlung natürlich entstandener Erkrankungen des Fesselträgers mit kontrollierter Bewegung allein oder kombiniert mit potenziell regenerativen Therapeutika . . . . .	24
Tab. 5.2.1	Verteilung der BehG-Anzahl, Behandlungsform (PrTB; KtrB) und Patienten-Fallzahl in Studien 1983-2011 und 2011-2016. . . . .	27
Tab. 5.2.2	Variabilität der Patienten-Fallzahl in den BehG . . . . .	28
Tab. 5.2.3	Verteilung der Patienten-Fallzahlklassen in den BehG . . . . .	29
Tab. 5.3.1	19 verschiedene prT in Studien 2001-2011 und 2011-2016 mit Angabe der Studien-Nummer . . . . .	30
Tab. 5.3.6	Absolute Anwendungshäufigkeit der 3 autologen, 4 allogenen und 2 xenogenen Gewebequellen in Studien 2001-2016 . . . . .	33
Tab. 5.4	Verteilung der 50 BehG (2001-2016) mit Angabe der <i>GFR</i> differenziert nach Behandlungsform und Gewebestruktur: OBS, TBS, FT sowie Kombinationen . . . . .	33
Tab. 5.4.2.2	11 Behandlungsgruppen-Paare der Behandlungsform PrTB . . . . .	37
Tab. 5.4.2.2.2	Ergebnis der Heterogenitätsanalyse zu PrTB: Bei 3 (der 11) BehG-Paaren können beide BehG gepoolt werden ( $42 \leq N-A \leq 145$ ) . . . . .	42
Tab. 5.4.3	4 alleinstehende BehG der Behandlungsform KtrB . . . . .	42
Tab. 5.4.3.1	Ergebnis der Heterogenitätsanalyse zu KtrB . . . . .	46
Tab. 5.4.4.1.1	'24 alleinstehende' BehG der Behandlungsform PrTB . . . . .	50
Tab. 5.4.4.1.1a	2 der '24 alleinstehenden' BehG zu PrTB unterscheiden die Gewebestruktur und haben die Fallzahl $30 \leq N-A$ . . . . .	52
Tab. 5.4.4.2a	Ergebnis der Untersuchung der BehG mit Angabe der Effektgröße <i>GFR</i> auf BehG, die sich möglicherweise für pV eignen; differenziert nach Behandlungsform (PrTB; KtrB), Gewebestruktur, Fallzahlklasse und Poolbarkeit . . . . .	52

Tab. 5.4.4.2b	Daten der 6 (von 31) BehG für mögliche paarweise Vergleiche (pV) zwecks Power-Analyse ( $30 \leq N-A \leq 145$ ) . . . . .	54
Tab. 5.4.4.3	Ergebnisse der Power-Analyse . . . . .	56
Tab. 5.6.3	Läsionsgrad (LG) nach RANTANEN et al. (2011) . . . . .	66
Tab. 5.7.3.1	Vergleich der Rehabilitationsdauer zu KtrB zwischen OBS-TBS vs. FT	70
Tab. 5.7.3.2	Vergleich der Rehabilitationsdauer zu PrTB zwischen OBS-TBS vs. FT	72
Tab. 5.7.4	Projektierte Rehabilitationsdauer gemäß <i>GFR</i> -Modell . . . . .	73
Tab. 5.7.8	Absolute Verteilung der Prüfzeit bei OBS-TBS vs. FT . . . . .	77
Tab. 5.7.9	Standardisierte Prüfzeit gemäß <i>GFR</i> -Modell bei OBS-TBS vs. FT	78
Tab 5.7.10	Standardisierte Studiendauer gemäß <i>GFR</i> -Modell . . . . .	78

# **Metaanalyse klinischer Studien 1983-2016 zur langfristigen Gebrauchsfähigkeit von Sportpferden nach Behandlung von natürlich entstandenen Erkrankungen der oberflächlichen und der tiefen Beugesehne und des Fesselträgers entweder allein mit kontrollierter Bewegung oder kombiniert mit einem potenziell regenerativen Therapeutikum**

## **1 EINLEITUNG**

Natürliche Erkrankungen des Sehnenapparates und Fesselträgers (FT) bei Sportpferden entstehen hauptsächlich durch kumulative Ermüdung oder mechanische Überbeanspruchung bei Training, Wettkampf oder Weidengang (GILLIS 1997). Am häufigsten betroffen sind die Gewebestrukturen der oberflächlichen (**OBS**) und tiefen (**TBS**) Beugesehne, das Unterstützungsband der tiefen Beugesehne (USB der TBS) sowie der Fesselträger (FT) (DYSON und GENOVESE 2003, DYSON 2007, MURRAY und DYSON 2006). Sehnenrisse sind bei Rennpferden häufig kein zufälliges Ereignis, sondern der Endpunkt eines chronischen Prozesses, der mit Mikrotraumata beginnt (POOL et al. 1990). Der kritische Bruchschwellenwert für Sehnenrisse beträgt beispielsweise bei der OBS ca. 16 % Längenausdehnung über normal (RICH und PATTERSON-KANE 2014). Beim FT wiederum sind unterschiedliche Gewebestrukturen betroffen. Der FT ist ein sehniger Muskel mit 2-11 % Muskelfaseranteil, im größeren Maß im Bereich der proximalen drei Viertel des FT-Körpers sowie einzelnen Muskelfaserbündeln im Bereich der Schenkel (SCHULZE und BUDRAS 2008). Durch andauernde Überdehnung des Fesselgelenks im Wettkampf beispielsweise bei Trabern wird der FT angespannt und ermüdet, was eine extreme Verletzungsgefahr birgt (JOHNSTON et al. 1999). Herkömmliche Therapieansätze führen nach einer langen Rekonvaleszenzphase zur Entstehung von Narbengewebe (reparatio, keine restitutio ad integrum) und infolge zu hohen Rezidivraten (GEBUREK und STADLER 2011b). Die wirtschaftlichen Folgeschäden sind erheblich.

Da Sehnenläsionen aufgrund der geringen regenerativen Kapazität des Gewebes nur unter Narbenbildung heilen, schlägt die equine Regenerationsmedizin mit einer ersten großen Studie (HERTHEL 2001; SMITH et al. 2003a) seit 2001 neue Wege ein. Intensiv wird im Labor in vitro, aber auch in vivo an künstlich erzeugten sowie natürlich entstandenen Sehnenverletzungen die Wirkung von potenziell regenerativen Therapeutika (**prT**) untersucht um die Bildung von Narbengewebe zu verhindern. Die positive Erwartungshaltung gegenüber prT scheint stetig zu wachsen. Gegenüber dem Zeitraum 2001-2011 (GEBUREK und STADLER 2011b) steigt in 2011-2016 die Zahl der durchschnittlich pro Jahr veröffentlichten klinischen Studien, die die Wirkung von prT bei natürlich entstandenen Erkrankungen des Sehnenapparates OBS, TBS und FT untersuchen, von 1,5 (17 /11) Studien pro Jahr auf 4,5 (27/6) um das Dreifache. Verletzungen der OBS im Bereich der metacarpalen Zone der equinen Gliedmaße, wie sie durch mechanische Überbeanspruchung beim Training oder Wettkampf entstehen, bieten unter Beachtung der großen individuellen Variabilität ein interessantes Tiermodell für Verletzungen der Achillessehne bei Sportlern, da sie die realen klinischen Bedingungen am Besten widerspiegeln (MUTTINI 2012; VERONESI et al. 2016).

PrT werden aus autologen, allogenen oder xenogenen Gewebequellen mit labortechnischen Herstellungsverfahren gewonnen. Die Zellen, Blutprodukte, biologischen Faktoren oder auch Gerüstsubstanzen (z. B. porcine 'urinary bladder matrix' UBM) werden intra- oder periläsional in das verletzte Sehnen- oder Bandgewebe appliziert. Sie sollen eine regenerative Heilung in Richtung einer Wiederherstellung des Gewebes fördern. Das Gewebe soll idealerweise die gleichen physiologischen Eigenschaften, insbesondere aber die gleiche biomechanische Leistungsfähigkeit haben wie vor der Verletzung. Einen Literatur-Überblick über die bisherigen Bemühungen bieten systematische Reviews (GEBUREK und STADLER 2011a und 2011b; VERONESI et al. 2016). Jedoch fehlt in der equinen Regenerationsmedizin der Nachweis des nachhaltigen langfristigen Behandlungserfolgs mit prT. Diese Dissertation möchte dazu beizutragen die bestehende Lücke zu

schließen. Methodisch eignet sich dafür gegenwärtig nur die Metaanalyse.

## 2 LITERATURÜBERSICHT

### 2.1 Definition von prT und therapeutisches Konzept

Der Oberbegriff 'potenziell regenerative Therapeutika' umfasst:

- 1) **Mesenchymale stromale multipotente Progenitorzellen (MSC)**. Sie werden aus perinatalem (z. B. Nabelschnurgewebe, Amniongewebe) oder postnatalem bzw. adultem Gewebe (z. B. Fettgewebe, Knochenmark) gewonnen (GEBUREK und STADLER 2011a; BURK und TROILLET 2016).
- 2) **Embryonale pluripotente Stammzellen**. Ihre Verwendung unterliegt ethisch-rechtlichen Vorschriften. Bemerkenswert ist, dass verletztes fetales Gewebe aufgrund seiner hohen regenerativen Kapazität ohne Bildung von Narbengewebe heilt (RIBITSCH et al. 2015).
- 3) **Blutprodukte**. BURK und TROILLET (2016) gliedern die Blutprodukte in blutzellbasierende prT aus Plasma und azelluläre prT aus Serum. Zu den blutzellbasierenden prT gehören: Autologes Plasma (**AP**); Thrombozytenkonzentrat (**PC**); Thrombozyten reiches Plasma (**PRP**, Platelet Rich Plasma); Thrombozyten reiches Filtrat (**PRS**, Platelet Rich Solution). Ein Serum basiertes prT ist autologes konditioniertes Serum (**ACS**). Daneben sind kommerzielle, in der Zusammensetzung vom Hersteller abhängige Kombinationspräparate als fertige Injektionslösungen erhältlich: z. B. Thrombozytenreiches Gel (**PRG**), Thrombozyten-Mediatoren-Konzentrat (**TMK**), Thrombozytenreiches Fibrin (**PRF**), Thrombozytenreiche Fibrin Matrix (**PRFM**) (GEBUREK und STADLER 2011a).

Das **therapeutische Konzept**, wonach die Sehnen und Bänderstrukturen nach Insult mit prT behandelt werden, stützt sich auf die Beobachtung, dass im verletzten Sehnengewebe verglichen mit der großräumigen Matrix nur wenige MSC verbleiben, was die Fähigkeit zur Heilung reduziert und den Vernarbungsprozess begünstigt (VANDENBERGHE et al. 2015). MSC können sich in verschiedene Zelllinien, Tendozyten, Chondroblasten, Osteoblasten der Stützgewebe differenzieren oder - nach neuerer Erkenntnis - deren Neubildung inszenieren oder zelluläre molekulare Reparaturvorgänge des Gewebes auslösen (ZWOLANEK et al. 2017). Patienten könnten daher deutlich davon profitieren, wenn zusätzlich zum natürlichen Heilungsprozess die passenden prT oder Wachstumsfaktoren gezielt verabreicht werden (GODWIN et al. 2012). Wie Erfolgsberichte aus der Humanmedizin nahelegen, kann so die ursprüngliche biomechanische Gewebefunktion stabil über mehrere Jahre zurückgewonnen werden (DOCHEVA et al. 2015).

Applizierte MSC verhindern die Aktivierung des angeborenen Immunsystem, indem sie die Monocyten- und T-Zell-Aktivität mittels signifikanter Reduktion von Tumornekrosefaktor-Alpha (TNF-alpha) und Interleukin (IL)-6 unterdrücken. Ebenso unterdrücken sie die B-Zell vermittelte humorale Immunantwort im Zielgewebe u. a. mittels verstärkter Produktion von anti-inflammatorischen Wirkstoffen wie IL-1 und IL-10 (BEERTS et al. 2013). Letzteres ermöglicht den Einsatz von MSC auch aus allogenen Gewebequellen (RYAN et al. 2005). Keine in vivo Studie der vorliegenden Metaanalyse (1983-2016) berichtet über Immunabwehrreaktionen. Nur wenige Fälle (z. B. bei UBM) mit jedoch beherrschbaren lokalen Entzündungsreaktionen werden berichtet; die Therapieform mit prT aus autologen wie allogenen Gewebequellen ist im allgemeinen biologisch verträglich (MUTTINI 2012; LANGE-CONSIGLIO et al. 2013); allerdings liegen noch keine systematischen Langzeitergebnisse vor (GEBUREK und STADLER 2011b).

## 2.2 Merkmale und Einflussfaktoren des langfristigen Behandlungserfolgs mit prT

### 2.2.1 Merkmale

Prinzipiell kann die Wirksamkeit einer Behandlung mit prT auf drei methodisch unterschiedlichen Ebenen untersucht werden:

- (1) Die biologische Qualität des applizierten prT; diese kann definiert werden nach Gewebequelle (ethische Bewertung) und regenerativem Differenzierungspotential (Inhaltsstoffe, biologische Verträglichkeit, therapeutische Verfügbarkeit, labortechnische Herstellungs- oder Expansionsverfahren, Verabreichungsform, Verabreichungsmenge etc.) (BURK und TROILETT 2016; VERONESI et al. 2016; GEBUREK et al. 2016).
- (2) Die feingewebliche, biomechanische und biochemische Qualität des Ersatzgewebes (GEBUREK und STADLER 2011a; VERONESI et al. 2016).
- (3) Die Qualität der GF der behandelten Sportpferde (GEBUREK und STADLER 2011b; VERONESI et al. 2016). Auf keiner Ebene liegen durch standardisierte quantitative Parameter oder standardisierte Protokolle gesicherte Erkenntnisse vor (BURK und TROILLET 2016; VERONESI et al. 2016).

Gegenwärtig mangelt es noch an standardisierten quantitativen Parametern zur objektiven Messung der strukturellen (Ultraschall) sowie ultrastrukturellen (Biopsie) feingeweblichen Eigenschaften des Ersatzgewebes (TETTA et al. 2012, BREHM et al. 2012, SMITH et al. 2014, DOCHEVA et al. 2015). Auch auf molekularbiologischer inter- und intrazellulärer Ebene fehlen für die potenziellen Einflussfaktoren und Effekte auf den Heilungsprozess des verletzten Sehnengewebe standardisierte Messgrößen (GASPAR et al. 2015). Es mangelt an evaluierten molekularen Nachweissystemen; viele Inhaltsstoffe sowie deren Interaktionen sind bisher unerforscht (BURK und TROILLET 2016). Noch können z. B. keine Genexpressionsprofile für die Sekretion chemotaktischer, bioaktiver oder immunmodulatorischer Faktoren über einen längeren Zeitraum hinweg erstellt, quantitativ zuverlässig gemessen und objektiv verglichen werden (DE SCHAUWER et al. 2013). Das resultierende Ersatzgewebe ist dem ursprünglichen gesunden Sehnengewebe – bei BM-MSC und Ad-MSC – qualitativ noch unterlegen (REED und LEAHY 2013). Obwohl neuere tierexperimentelle Studien mit prT andeuten, dass sie schneller zu einer Heilung führen und das resultierende Ersatzgewebe besser erscheint als das Narbengewebe, fehlt ein langfristiger Nachweis (GEBUREK et al. 2015). Läsionen, die in experimentellen Studien künstlich (chemisch; chirurgisch) erzeugt werden, sind keine gleichwertige Alternative zu Läsionen, die beim Wettkampf entstehen.

**Die vorliegende Metaanalyse entscheidet sich für Ebene 3.** Dadurch tritt die Auseinandersetzung mit Merkmalen, Einflussfaktoren und Messgrößen des Behandlungserfolgs auf der ersten und zweiten Ebene thematisch zurück. Der Vorteil liegt jedoch darin, dass sich die Unter- oder Überlegenheit der biologischen Qualität eines prT gegenüber einem anderen (Ebene 1), aber auch das Maß der Verbesserung der feingeweblichen Eigenschaften des resultierenden Ersatzgewebes im Vergleich zu Narbengewebe (Ebene 2) wahrscheinlich auf die Dauer und Intensität der Gebrauchsfähigkeit der behandelten Sportpferde (Ebene 3) auswirkt.

Auf Ebene 3 sollte die langfristige Überlegenheit einer regenerativen Behandlung mit 'prT kombiniert mit kontrollierter Bewegung' (**PrTB**) gegenüber einer konservativen mit 'kontrollierter Bewegung allein' (**KtrB**) an zwei Merkmalen erkennbar sein (GEBUREK und STADLER 2011b; HERTHEL 2001): (1) Die Rehabilitationsdauer bis zur Rückkehr zur Arbeit sollte kürzer sein. (2) Ein Rezidiv sollte signifikant weniger häufig auftreten.

**Ad (1)** Dieses Merkmal beschreibt den kurzfristigen Behandlungserfolg. Es gehört methodisch insoweit zu Ebene 3 als es langfristig bestätigt werden muss. Mit der Verkürzung der Heilungsdauer ist der Behandlungserfolg noch nicht nachhaltig gesichert (KÜNNEKE et al. 2008). Außerdem könnte es sich als ineffektiv erweisen, wenn dadurch das Risiko für ein Rezidiv steigt. Im Blick auf das Tierwohl wäre dies ethisch bedenklich.

Die Verbesserung des Sehnengewebes (Ebene 2) hat Priorität vor der Verkürzung der Rehabilitationsdauer (GODWIN et al. 2012). Eine Verkürzung der Rehabilitationsdauer durch prT ist gegenüber anderen Behandlungsformen zwar vereinzelt (z. B. ZUFFOVA et al. 2013 bei PRP), aber systematisch noch nicht evident nachgewiesen (KOCH et al. 2009; GOLETZ 2009; REED und LEAHY 2013). Das **GFR**-Modell bietet hierfür einen Lösungsansatz (5.8.1).

**Ad (2)** Es muss gezeigt werden, dass die Gebrauchsfähigkeit der behandelten Pferde langfristig rezidivfrei anhält. Das kann mit tierexperimentellen Studien nicht nachgewiesen werden. Allerdings ist bis jetzt kein standardisiertes valides Kontroll- und Bewertungsverfahren mit einer definierten quantitativen Effektgröße dafür verfügbar. Die Gründe liegen auf der Hand. Natürlicherweise stehen keine bezüglich der interessierenden Merkmale homogenen Untersuchungs- und Kontrollgruppen bereit um prT langfristig auf ihren Behandlungserfolg zu untersuchen. Außerdem verwenden, wie RICH et al. (2014) bemängeln, bisherige Studien unterschiedliche methodische Zugänge, u. a. nicht vergleichbare Erfolgsmerkmale: 'frei von Lahmheit', 'Rückkehr zur Nutzung auf dem ursprünglichen Leistungsniveau', 'Höhe der Siegesprämie'. Als Erfolgsmerkmale werden in den Studien zwar zwei quantitative Meßgrößen genannt: Erfolgsrate **E** (bzgl. Erfolgsmerkmal) und Rezidivrate **R**. Allerdings werden sie teils unterschiedlich berechnet. Oder sie fehlen in manchen Studien. Daher wird auf Ebene 3 in der vorliegenden Dissertation die Entwicklung eines evidenzbasierten Evaluationsmodells angestrebt, mit dem klinische in vivo Studien mit prT nach einem einheitlichen Standard geplant, durchgeführt und ausgewertet werden können.

### 2.2.2 Einflussfaktoren

In bisherigen Studien werden folgende potenzielle Einflussfaktoren auf den langfristigen Behandlungserfolg diskutiert: Gewebestruktur, Nutzungsrichtung bzw. sportliche Disziplin (Hindernisrennen, Flachrennen, Dressur, Springen, Ausdauer etc.) sowie Grad, Art und Alter der Läsion bzw. der Erkrankung bei Behandlungsbeginn und das Alter der Sportpferde.

**Gewebestruktur:** Sehnengewebe OBS, TBS einerseits und FT andererseits mit den drei Segmenten Ursprung, Körper, Schenkel, unterscheiden sich in der physiologischen Gewebequalität sowie biomechanischen Beanspruchung bei verschiedenen sportlichen Disziplinen; daher scheint es geboten, die therapeutische Wirkung von prT nach der Gewebestruktur zu differenzieren (GEBUREK und STADLER 2011b; BURK und BREHM 2011).

**Nutzungsrichtung:** Wie bereits angedeutet unterscheiden sich die physiologischen Vorgänge bei OBS, TBS sowie FT bei spezifischen Trainings-Formen, Trainings-Geschwindigkeiten, Gangarten (Schritt, Trab, Galopp) und sportlichen Leistungsstufen entsprechend der biomechanischen Arbeitsbelastung. Das Gewicht von Pferd und Reiter verteilt sich während des Bewegungsablaufs beim Ausüben einer bestimmten sportlichen Disziplin funktionell bedingt unterschiedlich auf den Bewegungsapparat von Vorder- und Hintergliedmaßen (TAKAHASHI et al. 2004; TRUMP 2014; SMITH 2016; BANSE 2011; LUTZ 2011). MARNERIS und DYSON (2014) beobachten an 71 Sportpferden, dass bspw. 76,2 % der Eventpferde Verletzungen der FT-Schenkel verstärkt an den Vorder-, dagegen 87,1 % der Dressurpferde an den Hintergliedmaßen haben. Insofern beeinflusst die Nutzungsrichtung Entstehung und Heilung entsprechender Erkrankungen, wie auch wahrscheinlich das Auftreten eines Rezidivs. Zwischen unterschiedlichen Disziplinen streut die Rezidivrate **R** erheblich (BURK und BREHM 2011). Eine bestimmte sportliche Disziplin, der Pferdetyp (Rasse: Warmblut, Kaltblut, Araber, Thoroughbred, Standardbred etc.) und die Verletzungsform passen häufig zusammen. So verletzen sich speziell für den Rennsport gezüchtete Thoroughbreds vermehrt an der OBS, da sich bei schnellen Bewegungsabläufen die Gewebetemperatur zentral stark erhöht und Mikro-Einrisse an Sehnenfasern entstehen, die zwar reparativ verheilen, aber durch stetige Wiederholung des Vorgangs zu größeren Risiken führen (SMITH et al. 2009). Sie bilden typisch hohlraumartige 'core lesions' mit sonographisch hypo-



echogenem Zentrum und scharf abgegrenzten Rändern aus intakten Sehnenfasern (PATTERSON-KANE und FIRTH 2009). Cutting-Horses dagegen zeigen oft schon im jungen Alter typisch periphere laterale Sehnenverletzungen leichten bis mittleren Grades mit kleinen CSA, die aber sonographisch nicht scharf, sondern diffus berandet erscheinen. Mit KtrB behandelte Cutting-Horses kehren meist nach einer mittleren Rehabilitationsdauer von 4 Monaten zum vollen Training zurück (TIPTON et al. 2013). Werden an die Nutzungsrichtung angepasste Trainingsübungen in das kontrollierte Bewegungsprogramm integriert, kann dies den Heilungsprozess in der Remodellierungsphase positiv beeinflussen. GILLIS (1997), RANTANEN et al. (2011) und SMITH (2016) legen detaillierte Anleitungen zum kontrollierten Bewegungsaufbautraining vor (vgl. Exkurs).

**Läsionsgrad.** Einzelnen Studien zufolge beeinflusst der Läsionsgrad sehr wahrscheinlich die Chance der erfolgreichen Rehabilitation und Rückkehr zur Arbeit (YOVICH et al. 1995). Der Einfluss ist jedoch in der Literatur noch nicht systematisch untersucht (WITTE et al. 2011).

**Alter der Pferde.** Das Alter erscheint als ein wichtiger Einflussfaktor. Im Zusammenhang mit Gewebestruktur, Nutzungsart und Läsionsgrad scheint er eine limitierende Rolle zu spielen. Nach KASASHIMA et al. (2004) ist das Risiko für eine Verletzung der OBS bei älter als fünfjährigen Flachrennpferden (ca. 16 %) drei Mal so hoch wie bei zweijährigen (ca. 6 %). Generell verlieren im Alter die Sehnen ihre Fähigkeit sich selbst zu erneuern; Bindegewebe wird allgemein steifer und ist weniger reißfest. Die histomorphologische Untersuchung belegt degenerative Veränderungen der Tenozyten und einen Strukturverlust der extrazellulären Sehnenmatrix (THORPE et al. 2010). Die zunehmende Anzahl von DNA-Schäden im Laufe des Lebens führt zu einer reduzierten reparativen Kapazität und Remodellierung der Kollagen I Sehnenmatrix. Bisherige Studien gehen von einer günstigeren Prognose für jüngere Pferde aus bei Behandlung mit MSC (GODWIN et al. 2012). Spätestens seit KERN et al. (2006) ist bei Knochenmark als Gewebequelle bekannt, dass auch alterungsbedingte degenerative Sehnenerkrankungen des Donors der MSC beim Empfänger die Heilung nach Applikation negativ beeinflussen können. Dies gilt auch für Spender mit Fibroplasie (BURK 2012). Bei Sehnenzellen deuten Mikro-Array und PCR Daten in vitro beim Vergleich zwischen jungen und alten Spendern von MSC darauf, dass die Aktin-Dynamik im Zytoskelett gestört ist, was einhergeht mit einer Dysregulation des Integrin-Rezeptor-Profiles (DOCHEVA et al. 2014).

**Geschlecht.** Das Geschlecht scheint als direkter Einflussfaktor eine untergeordnete Rolle zu spielen (PLEVIN und MCLELLAN 2014); jedoch ist bekannt, dass das höhere Gewicht männlicher Rennpferde ein erhöhter Risikofaktor für Tendinopathien sein kann (TAKAHASHI et al. 2004).

**Zwischenbilanz und Schlussfolgerung.** Will man auf Ebene 3 mithilfe der metanaalytischen Methode in der Frage einen Schritt vorankommen, ob sich die 'langfristige Gebrauchsfähigkeit' (GF) der Sportpferde zwischen einer regenerativen und konservativen Behandlung oder zwischen verschiedenen prT unterscheidet, so erscheint es methodisch unumgänglich ein Evaluationsmodell zu entwickeln, mit dem der Behandlungserfolg mittels *E* und *R* quantitativ valide gemessen werden kann. Dazu ist (1) das Erfolgsmerkmal (einheitlich) zu definieren; (2) die Rehabilitationsdauer zu messen; (3) die Dauer der Prüfzeit zu standardisieren, (4) *E* und *R* normiert zu berechnen und (5) zu einer einzigen (vergleichbaren) quantitativen Effektgröße zusammenzufassen. Folgende Primärdaten sollten als maßgebliche Einflussfaktoren definiert werden: (1) Gewebestruktur (OBS, TBS, USB der TBS, FT-Ursprung, FT-Körper, FT-Schenkel) und Lokalisation (vorne, hinten) der Gliedmaße (GEBUREK und STADLER 2011b; BURK und BREHM 2011); (2) Nutzungsrichtung (GEBUREK und STADLER 2011b; BURK und BREHM 2011); (3) Läsionsgrad und Alter der Läsion bei Behandlung (PLEVIN und MCLELLAN 2014; ZUFFOVA et al. 2013); (4) Altersklasse der Pferde (KASASHIMA et al. 2002; ROSENBROCK et al. 2004B; JORGENSEN et al. 2011; BURK und BREHM 2011; SCALA et al. 2014).

## 2.3 Die Metaanalyse als Modell zur Bewertung klinischer Studien

### 2.3.1 Die Metaanalyse als Methode der evidenzbasierten Veterinärmedizin

Der Begriff 'Evidenz' bezeichnet im Deutschen 'Einsichtigkeit' (lat. Evidentia, abgeleitet von videre, sehen). Im englischen Sprachraum bezeichnet 'evidence' Beleg, Beweis. Beides erscheint auf den ersten Blick gegensätzlich, denn für das unmittelbar Einsichtige bedarf es keines Beweises. Jedoch unterscheidet sich nur die Blickrichtung; in beiden Fällen ist die Grundbedeutung das 'Sehen' (videre) als (wissenschaftliche) Suche nach der wahren Antwort auf eine einfache Frage. Bei der Bedeutung 'Einsichtigkeit' richtet sich der Blick auf das Ergebnis der Suche, die einsichtige Antwort; dagegen bei 'Beweis' auf den Vorgang (den Weg, Beweisgang) der Suche.

Das Konzept der evidenzbasierten Veterinärmedizin **EBVM** hat – aus historischer und wissenschaftstheoretischer Sicht – Malynicz (MALYNICZ 1998) angeregt, der dabei an die von David Sackett (SACKETT et al. 1996) entwickelte evidenzbasierte Medizin **EBM** anschließt. EBVM und EBM fordern den Gebrauch der aktuell besten Evidenz klinischer Erkenntnisse und wissenschaftlicher Forschungsergebnisse bei der medizinischen Behandlung und Versorgung von Patienten; dabei bedeutet Evidenz die Sicherheit, dass die Erkenntnisse und Ergebnisse tatsächlich richtig sind (ARLT 2013). In den letzten beiden Jahrzehnten arbeiten vor allem die Cochrane Collaboration, das Deutsche Netzwerk evidenzbasierte Medizin e.V., die Evidence-based Veterinary Medicine Association und das Centre for Evidence-based Veterinary Medicine, daran, die Ziele der EBM bzw. EBVM in Praxis, Forschung und Ausbildung zu verwirklichen.

Das systematische Review und die Metaanalyse sind zwei Studiendesigns, die für EBM und EBVM eine Schlüsselrolle einnehmen. Ersteres analysiert zu einer einfachen Frage die relevanten Informationen einer repräsentativen Zahl von Forschungsstudien, vergleicht die qualitativen und quantitativen Ergebnisse und kommt zu einer qualitativen Schlussfolgerung, d. h. zu der bestmöglichen Beantwortung der Frage (HIGGINS und GREEN 2011; ZIEGLER et al. 2007). Die Studien-Designs der untersuchten Studien werden gemäß Evidenz-Stufenmodell hierarchischen geordnet (ARLT 2013). Das Stufenmodell zeigt planvoll auf, wie durch das Studiendesign bei der Durchführung einer Studie sowie Analyse der Studienergebnisse und Berichterstattung systematische Fehler (Bias) reduziert werden können (ARLT 2013). Deskriptive Fallstudien rangieren auf der niedrigsten, metaanalytische Studien von randomisierten kontrollierten Studien auf der höchsten Evidenzstufe (ARLT 2013). Erstere deshalb, da eine Kontrollgruppe fehlt und sich so nicht ermitteln lässt, welcher Faktor einer therapeutischen Intervention tatsächlich einen positiven oder negativen Effekt hat. SARGEANT et al. (2014) entwerfen eine Evidenz-Pyramide der Studien-Designs, wonach in Verbindung mit Randomisierung natürliche Studienbedingungen im Gegensatz zu künstlichen Laborbedingungen zu einer steigenden Evidenz beitragen. Der Evidenz-Grad richtet sich nach dem Maß an Validität der qualitativen und quantitativen Ergebnisse; das ist das Maß an Freiheit der Messergebnisse und Schlussfolgerungen von Bias (ARLT 2013). Bias entsteht bei irgendeinem Prozess in der Beweisführung zu irgendeinem Zeitpunkt und führt dazu, dass Resultate produziert werden, die sich systematisch vom tatsächlichen Wert unterscheiden (MURPHY 1976). Das Maß an Freiheit von Bias ist um so größer, je besser die Reduktion systematischer Fehler gelingt (JÜNI et al. 2001). Dabei bezieht sich die interne Evidenz einer Studie auf das Maß, in dem die quantitativen Ergebnisse und qualitativen Schlussfolgerungen dem wahren Sachverhalt entsprechen (ARLT 2013), was den höchstmöglichen Grad der Freiheit von systematischen Fehlern einschließt (PERLETH und RASPE 2002). Die externe Evidenz dagegen sagt etwas aus über das Maß, mit dem die Studienergebnisse für die klinisch-therapeutische Anwendung verallgemeinert werden können (KHAN et al. 2001). Insofern bedingen auf höchster Evidenzstufe beide einander (JÜNI et al. 2001).

Eine zweite Schlüsselkomponente der EBM ist die Metaanalyse. Sie wird als zentrales Instrument der medizinischen Forschung betrachtet, um evidenten Forschungswissen in die Praxis zu übertragen und wichtige

Fragestellungen in der Forschung voranzutreiben (PETRISOR und BHANDARI 2006; MILES et al. 2010). Sie geht bei der Beantwortung einer einfachen Frage wie das systematische Review vor, versucht jedoch darüber hinaus die quantitativen Ergebnisse (positive wie negative) aller Studien in einem quantitativen Ergebnis zu integrieren. Die erste Metaanalyse geht auf den Statistiker Karl Pearson im Jahr 1904 zurück (PEARSONS 1904). Er erkennt, dass die Anzahl der Probanden in Studien oft zu gering ist um daraus statistisch gesicherte quantitative Schlussfolgerungen für die Gesamtpopulation zu ziehen. Daher fasst er zu einer bestimmten Frage die quantitativen Primärdaten mehrerer inhaltlich homogener Studien zusammen. Allerdings folgen in den Jahrzehnten danach nur wenige Studien dem Weg Pearsons. Zu einem breiten Aufschwung seines Grundgedankens kommt es erst in den 1970er-Jahren im Zuge der 'evidence movement' (HANSEN und RIEPER 2009), in deren Verlauf der Statistiker und Psychologe Gene Glass den Begriff 'metaanalysis' prägt (GLASS 1976; CHALMERS et al. 2010). Aus der 'evidence movement' geht die EBM hervor. Mit zunehmender Leistungsfähigkeit digitaler Rechenmaschinen seit den 80er-Jahren entstehen geeignete Softwaretools zur statistischen Analyse, Verarbeitung und metaanalytischen Integration auch großer quantitativer Datenmengen homogener Studien. Im Falle des *GFR*-Modells kann mithilfe von statistischen Rechenprogrammen der Heterogenitätsanalyse z. B. geprüft werden, ob bei der Integration der quantitativen Effektgröße mehrerer homogener Behandlungsgruppen (BehG), die aus den Studien extrahiert werden, Unterschiede der Varianz zwischen den BehG mittels quantitativer kumulativer Gewichtung ausgeglichen werden können oder ob im Falle von Heterogenität Anteile der Varianz (Residuen) verbleiben, die wiederum auf Moderatoren (Drittvariable) hinweisen. Allgemein können letztere etwa durch die wissenschaftliche Fragestellung, die Methode (Messung, Studiendesign etc.) oder die Qualität der Primärdaten bedingt sein. Mittels Robustheit der Effektgröße gegenüber einer Auswahl von Einflussvariablen können einerseits stabile Beziehungen zwischen den Variablen detektiert, andererseits mögliche unterschiedlich starke Auswirkungen auf den Behandlungseffekt einer Population erkannt werden. Die abschließende qualitative Gesamtbewertung stellt den höheren Informationsgehalt der Analyseergebnisse vor oder generiert neue Hypothesen. Dieser abschließende Schritt ist notwendig, da die Ergebnisse von Metaanalysen immer Zahlen sind und daher der angemessenen Interpretation für die veterinärmedizinische Praxis bedürfen (STAMM und SCHWARB 1995). Die geeigneten statistischen Methoden werden in jüngerer Zeit etwa von TIMMER und RÜCKER (2008), RESSING et al. (2009), LEAN et al. (2009), ODGAARD-JENSEN et al. (2011), SIMONEIT et al. (2012), ZIEGLER et al. (2007), PREUSS (2013) unter Rückgriff auf die zugrundeliegende mathematische Fachliteratur beschrieben.

### 2.3.2 Drei Typen der Metaanalyse

Die 'evidence movement' entwickelt drei Typen der Metaanalyse: (1) Die Metaanalyse publizierter Daten; (2) die gepoolte Reanalyse; (3) die prospektiv geplante Metaanalyse (BLETTNER et al. 1997; RESSING et al. 2009; ARLT 2013).

**Ad (1)** Die 'Metaanalyse publizierter Daten' ist ein retrospektives Studiendesign. Zu einer einfachen Frage werden die Kriterien für die Auswahl der Primärstudien festgelegt. Die Forschungsergebnisse möglichst aller zu der Fragestellung inhaltlich homogenen Studien werden zusammengetragen. Die quantitativen Ergebnissen der Primärstudien werden als Stichprobe betrachtet und darauf die geeigneten deskriptiven und induktiven Methoden der mathematischen Statistik angewandt; das Ergebnis der Schätzung der Stichprobenparameter wird abschließend qualitativ bewertet (SCHMIDT 2007). Die Effektgrößen aller Studien müssen vergleichbar oder in solche transformierbar sein um mit statistischen Methoden in ein quantitatives Gesamtergebnis integriert werden zu können; möglicherweise sind Subgruppen zu bilden. Nicht bei jeder einfachen Frage ist die Voraussetzung der Vergleichbarkeit, Transformierbarkeit, Subgruppenbildung etc. erfüllt; daher ist eventuell ein Systematisches Review, aber nicht notwendig eine Metaanalyse dazu durchführbar (LEAN

et al. 2009). Beispiele dafür aus jüngster Zeit finden sich etwa in der Studie von WAGNER (2014) „Meta-analyse zur Therapie der bovinen Dermatitis digitalis“.

**Ad (2)** Die 'gepoolte Reanalyse' unterscheidet sich von Typ (1) methodisch darin, dass sie auf die Quelldaten der Primärstudien (kurz „Primärdaten“) zurückgreift, diese analysiert, mit geeigneten statistischen Methoden bearbeitet und zusammenfasst, d. h. poolt. Die Quelldaten aller Studien werden dabei – unabhängig von der Bewertung der Primärstudien – neu bewertet. Neuauswertungen gepoolter metaanalytischer Daten sind zwar wesentlich zeitaufwendiger oder kostenintensiver als Metaanalysen publizierter Daten, erhöhen jedoch die Chance auf zuverlässige Ergebnisse (BLETTNER und SAUERBREI 1998; PREUSS 2013). Die vorliegende Metanalyse mit einem neu entwickelten evidenzbasierten Evaluationsmodell ist hier einzuordnen.

**Ad (3)** Die 'prospektiv geplante Metaanalyse' wird im Blick auf die prognostizierten Auswirkungen der Intervention geplant. Mit zukünftig durchzuführenden klinischen Studien sollen – etwa in einem Forschungsverbund – zu einer einfachen Fragestellung von Anfang an vergleichbare Primärdaten erzeugt, anhand einer definierten Effektgröße statistisch ausgewertet und das Ergebnis der Schätzung qualitativ bewertet werden. Das kann analog auf Studien mit prT übertragen werden; denn das hier vorgestellte Evaluationsmodell ermöglicht klinische Studien unter gleichen kontrollierten Bedingungen.

### 2.3.3 Die Metaanalyse in der equinen Regenerationsmedizin

**Ausgangslage.** Seit 1983 liegen für die Behandlungsform PrTB nur 3 RCTs vor. Zwei RCTs zur Behandlung mit PRP: GARETT et al. (2013), n = 20, beim FT-Schenkel; GEBUREK et al. (2016), n = 20, bei der OBS. Und eine RCT zur Behandlung mit ACS: GEBUREK et al. (2015), n = 15, bei der OBS; die Kontrollgruppe (n= 5) behandelt mit KtrB. Vielversprechende klinische Behandlungsergebnisse überwiegend unkontrollierter klinischer Studien liegen dagegen zahlreich vor; seit 1983 ca. 51 Studien. Allerdings sind die Ergebnisse nicht oder bedingt vergleichbar. Als Folge wird gegenwärtig noch kontrovers diskutiert, ob eine Therapie mit MSC aus Fettgewebe einer Therapie mit MSC aus Knochenmark unter- (FRISBIE und SMITH 2010) oder überlegen (RICH et al. 2014; GUERCIO et al. 2015) ist; biologische Unterschiede deuten sich an, die wahrscheinlich Konsequenzen für die therapeutische Wirkung haben (MARX et al. 2015). Ad-MSC können sicherer und leichter isoliert werden und bieten die größere Anzahl an frisch verfügbaren MSC (GUERCIO et al. 2015). Da jedoch generell offen bleibt, ob sich der Erfolg der regenerativen Behandlung zwischen verschiedenen prT unterscheidet, beruht bislang die Entscheidung für das eine oder andere vielversprechende Präparat nur auf empirischer Evidenz (VERONESI et al. 2016). Einer aktuellen Umfrage zufolge halten sich Pferdebesitzer in Australien bei equinen Sehnen- und Bänderkrankungen mehrheitlich von einer regenerativen Behandlung mit prT zurück, da die nachhaltige Gebrauchsfähigkeit der Pferde noch nicht evident erwiesen ist (MARTIN et al. 2014). Angesichts der ungelösten Fragen erscheinen konkrete Ergebnisse für die therapeutische Anwendung bei Sportpferden um so dringender. Immer wieder wird daher eine koordinierte Zusammenarbeit innerhalb der veterinärmedizinischen Forschungsgemeinde angemahnt (DE SCHAUER et al. 2013, ARLT 2013). Gegenwärtig scheint nur die Metaanalyse geeignete Methoden anzubieten um hierbei Fortschritte zu erzielen.

Der bisher einzige metaanalytische Versuch zur equinen Regenerationsmedizin ist die Studie „Regenerative Therapie von Sehnen- und Bänderkrankungen bei Pferden: Ergebnisse der Behandlung mit Stammzellen, Blutprodukten, Gerüstsubstanzen und Wachstumsfaktoren – eine Literaturübersicht und Metaanalyse“ von GEBUREK und STADLER (2011b). Die Literaturübersicht stellt die bisherigen Erkenntnisse zum regenerativen Potenzial von MSC sowie Blutprodukten vor. Es liegen 3 Studiendesigns vor: 1) Klinische Studien, deren Ergebnisse zumeist mithilfe einer Vergleichsgruppe kontrolliert werden; 2) tierexperimentelle kontrollierte Untersuchungen an künstlich erzeugten Sehnendefekten mit wertvollen Ergebnissen zu feingeweblichen, biochemischen und biomechanischen Eigenschaften des Ersatzgewebes; 3) einzelne plazebokontrollierte Un-

tersuchungen an natürlich entstandenen Tendinopathien. Insbesondere präsentiert der metaanalytische Teil die Ergebnisse von 21 klinischen Studien (1983-2011) zur Behandlung natürlich entstandener Erkrankungen des Sehnenapparates OBS, TBS und FT, mit KtrB oder PrTB. Allerdings haben die meisten Studien eine methodische Schwäche, so dass ein wesentliches Ziel, die unterschiedlichen Behandlungserfolge zu vergleichen, nicht erreicht werden kann. 67 % (14 /21) belegen den Behandlungserfolg nach Behandlung und erfolgreicher Rehabilitation zwar anhand der Erfolgsrate  $E$  des Erfolgsmerkmals (EM) 'Rückkehr zur Nutzung auf dem ursprünglichem Leistungsniveau', kontrollieren jedoch nicht über einen längeren Prüfzeitraum mittels Rezidivrate  $R$  wie lange die Gebrauchsfähigkeit anhält. Selbst die Ergebnisse der restlichen 7 Studien sind nur bedingt vergleichbar, da sie auf unterschiedlich langen Prüfzeiträumen für  $R$  basieren. Der Prozentsatz der Studien, die beide Anteilsraten,  $E$  und  $R$ , angeben oder aus ihren Primärdaten zu schätzen erlauben, verdoppelt sich im Zeitraum 2011-2016 gegenüber 1983-2011 von 33 % (7/21) auf 66 % (20/30). Allerdings beziehen viele Studien  $E$  und  $R$  auf Kombinationen von Gewebestrukturen. Der Ansatz ist nicht zielführend, da nicht willkürlich davon ausgegangen werden kann, dass ein und dasselbe prT bei allen Gewebestrukturen therapeutisch gleichermaßen wirkt. Die Ergebnisse sind nicht vergleichbar. Vielmehr müssen die physiologisch strukturellen und funktionell-biomechanischen Unterschiede von Sehnen- und Bändergewebe zwischen OBS, TBS, FT-Ursprung, FT- Körper, FT-Schenkel, distale Insertion des FT an den Gleichbeinen berücksichtigt werden. Auch bleibt offen, wie  $E$  und  $R$  quantitativ verknüpft (5.5.3) werden kann.

### 3 FRAGESTELLUNG, ZIELE UND HYPOTHESEN

Fragestellung und Zielsetzung der Dissertation orientieren sich an folgender Ausgangslage: Natürlich entstandene Erkrankungen des Sehnenapparates bei Sportpferden betreffen am häufigsten oberflächliche (OBS), tiefe (TBS) Beugesehne sowie Fesselträger (FT). Konventionelle Behandlungsformen führen zu konsekutivem Narbengewebe, woraus hohe Rezidivraten ( $R$ ) resultieren. PrT sind mesenchymale stromale multipotente Progenitorzellen (MSC) aus unterschiedlichen Gewebequellen, sowie blutzellbasierende (meist PRP, Platelet Rich Plasma) oder azelluläre Blutprodukte. Die equine Regenerationsmedizin steht gegenwärtig vor der Frage, ob eine regenerative Behandlung mit 'prT kombiniert mit kontrollierter Bewegung' (PrTB) einer konservativen mit 'kontrollierter Bewegung allein' (KtrB) bzgl. der 'langfristigen Gebrauchsfähigkeit' (kurz: GF) überlegen ist. Allerdings gibt es für GF bislang keine quantitative Effektgröße. Auch sind die Ergebnisse bisheriger Studien zum Behandlungserfolg von prT nicht vergleichbar. Häufig sind Erfolgsmerkmal (EM) oder Erfolgsrate  $E$  (bzgl. EM) unterschiedlich definiert; oder ist die Prüfzeit ( $P$ ) zur Kontrolle von EM auf ein Rezidiv und zur Bestimmung von  $R$  unterschiedlich lang. Oft wird  $R$  unterschiedlich berechnet oder fehlt. Oder es werden unterschiedliche Gewebestrukturen kombiniert und nicht differenziert untersucht. Eine systematische Klärung der Frage des langfristigen Behandlungserfolgs mit prT erscheint gegenwärtig nur mithilfe der metaanalytischen Methode möglich und ist das Ziel dieser Dissertation.

Die zielführenden Arbeitsfragen lauten: Wie kann mit einer einzigen Effektvariable die GF von Sportpferden, die nach einer natürlich entstandenen Erkrankung der OBS, TBS oder FT mit KtrB oder PrTB behandelt werden, differenziert nach der Gewebestruktur quantitativ valide gemessen werden? Und wie können die bisherigen Studienergebnisse verglichen werden?

Folgende Hypothesen werden aufgestellt:

- (1) Die metaanalytische 'gepoolte Reanalyse' kann mithilfe eines evidenzbasierten einfachen Evaluationsmodells mit der neuen quantitativen Effektgröße 'Gebrauchsfähigkeitsrate' **GFR** die Ergebnisse der bisherigen Studien (1983-2016) zu PrTB und KtrB unabhängig von den Ergebnissen der Primärstudien neu bewerten und vergleichen. **GFR** kann aus  $E$  (bzgl. EM) und  $R$  berechnet werden. EM lautet: 'Nach Rehabilitation kehren die Sportpferde zur Arbeit auf (mindestens) dem ursprünglichen Leistungsniveau zurück'. Mittels  $R$  wird kontrolliert, welcher Anteil der Sportpferde in der standardisierten Prüfzeit rezidivfrei bleibt.
- (2) Aus den Studien können kleine Behandlungsgruppen (**BehG**) mit gleichen interessierenden Merkmalen differenziert nach Gewebestruktur (OBS, TBS, USB der TBS, FT-Ursprung/-Körper/-Schenkel) und Behandlungsform (KtrB, PrTB) extrahiert und daraus **GFR** bestimmt werden.
- (3) Die Berechnung von **GFR** aus nur 3 Variablen  $A$ ,  $B$ ,  $C$  kann normiert und die Studiendauer im Konsens mit der Forschung standardisiert werden. Da sie sich aus der projektierten Rehabilitationsdauer (**proj. Reha**) und der Prüfzeit ( $P$ ) summiert, sind beide zu standardisieren. Aufgrund der gleichen Messbedingungen ist **GFR** zwischen BehG vergleichbar.
- (4) Kleine inhaltlich homogene BehG können auf Bias geprüft, ihre **GFR** mittels quantitativer Heterogenitätsanalyse gepoolt und der resultierende Effektschätzer  $GFR_{\text{gepoolt}}$  (mit Konfidenzintervall) berechnet werden. Unterschiede der **GFR** zwischen BehG können bei ausreichender Fallzahl per Signifikanznachweis mittels Power-Analyse statistisch gesichert werden. **GFR** quantifiziert valide die rezidivfreie GF von Sportpferden nach Behandlung mit PrTB oder KtrB für 1,5 Jahre beim FT ( $1,5GFR$ ) bzw. für 2 Jahre ( $2GFR$ ) bei OBS, TBS bei natürlich entstandenen Erkrankungen.
- (5) Zukünftige Studien gemäß **GFR**-Modell können replizierbare Ergebnisse von hoher interner Evidenz generieren. Werden sie vergleichbar dem Konzept prospektiv geplanter Metaanalysen durchgeführt, besteht die Chance auf Ergebnisse von hoher externer Evidenz.

## 4 MATERIAL UND METHODEN

### 4.1 Literatursuche

#### 4.1.1 Ein- und Ausschlusskriterien

Gesucht werden klinische In-vivo-Studien im Zeitraum 2011-2016, die natürlich entstandene Erkrankungen des Sehnenapparates OBS, TBS und FT bei Sportpferden mit KtrB oder PrTB behandeln. Dabei gelten folgende Ein- bzw. Ausschlusskriterien, die hier aus Gründen der Übersichtlichkeit kurz tabellarisch zusammengestellt werden:

Tabelle 4.1.1	Einschlusskriterien	Ausschlusskriterien
	natürlich entstandene Sehnen- bzw. FT-Erkrankungen	induzierte Sehnen- bzw. FT-Erkrankungen
	Verwendung von kontrollierter Bewegung allein oder kombiniert mit einem prT	Verwendung von chirurgischen oder konventionellen Behandlungsmethoden ohne prT
	Sprachen: deutsch, englisch, spanisch, italienisch	andere Fremdsprachen

#### 4.1.2 Online-Recherche

Die Fachartikel sind vorwiegend online über das VPN-Netzwerk der Universitätsbibliothek Leipzig recherchiert. Grundlage der Literatursuche sind die Datenbanken PubMed (<http://www.pubmed.gov>) und CAB Abstracts ([www.cabdirect.org](http://www.cabdirect.org)). Erstere gehört der National Library of Medicine (NLM) an und stellt die weltweit größte elektronisch verfügbare medizinische Datenbank dar. Sie verfügt nur über englischsprachige Literatur. Die Referenzdatenbank CAB Abstracts wird von CABI (Commonwealth Agricultural Bureaux International) unterhalten und beinhaltet neben zahlreichen Fachgebieten die wichtigste Spezialdatenbank der Veterinärmedizin. Sie enthält bibliographische Einträge und englische Abstracts zu Veröffentlichungen aus mehr als 120 Ländern in 50 Sprachen. Die Sprachauswahl der Suche ist auf Englisch, Deutsch, Spanisch und Italienisch eingegrenzt. Seit 2003 kommen bei CAB auch PDF Volltext-Artikel hinzu: Wissenschaftliche Journale, Buchkapitel, Monographien, Regierungsberichte, Konferenzbeiträge (oft vor Veröffentlichung der vollständigen Studien) mit jährlich ca. 10.000 neuen Titeln. Daraus stammen eine spanische (ABBELANET et al. 2009) und eine italienische (TASSAN et al. 2014) klinische Studie.

#### 4.1.3 Manuelle Recherche

Nicht im Internet verfügbare Fachartikel und Kongressbände sind aus dem Präsenz-Bestand der veterinärmedizinischen Bibliothek der Universität Wien kopiert oder über den kostenpflichtigen Dokumentenlieferdienst [www.subito-doc.de](http://www.subito-doc.de) bestellt. Nach dem Prinzip der „Schneeball-Literatursuche“ sind die Literaturverzeichnisse aller bereits gefundenen Artikel auf zielführende Literatur ausgewertet. Aus dem Gespräch mit Kollegen auf Fachkongressen stammen Hinweise für weitere thematisch verwertbare Studien, ebenso aus der Durchsicht von Kongressbänden.

#### 4.1.4 Suchstrategien in der Datenbank PubMed

Die Literatursuche in der Datenbank PubMed basiert auf hierarchisch geordneten Schlagwörtern, sog. MeSH-Terms (Medical Subject Headings). d. h. in die Suche nach einem übergeordneten MeSH-Term sind unter-

geordnete inhaltlich zusammengehörige Schlagwörter eingeschlossen. Schlagwörter können mit den Operatoren der booleschen Algebra - *und* (die Suche einengend), *oder* (die Suche erweiternd), *nicht* (die Suche ausschließend) - logisch verknüpft werden (DICTY 2007). Da es thematisch um Erkrankungen des Sehnenapparates und FT bei Sportpferden geht, sind *horses* und *tendons* übergeordnete MeSH-Terms. Sie werden wie folgt schrittweise mit spezielleren Suchbegriffen verknüpft. Von besonderem Interesse sind diejenigen Sehnen und Bänder der Pferdegliedmaße, die am häufigsten von Läsionen betroffen sind: Die oberflächliche /tiefe Beugesehne, *superficial digital flexor tendon* /*deep digital flexor tendon* und der FT, *suspensory ligament*. Sekundär interessieren die Unterstützungsbänder der oberflächlichen /tiefen Beugesehne (*accessory ligament of the ... flexor tendon*), deren Erkrankungen in wenigen Studien mitbehandelt werden.

Die gesuchten Studien sollen Informationen darüber enthalten, wie viele Patienten nach der Behandlung das Erfolgsmerkmal erreichen, d. h. zum ursprünglichen bzw. höheren Leistungsniveau zurückkehren oder rekonvaleszent 'reconvalescence' bleiben. Für das Erfolgsmerkmal sind in den klinischen Studien folgende 'return to'-Formulierungen gängig: *return to work*, *...full work*, *...activity*, *...use*, *...race*, *...racing*, *...previous level*, *...sport*. Des weiteren interessiert, ob die Patienten nach der Rehabilitation 'to be rehabilitated' in der Nach-Behandlungsphase (follow up) ein Rezidiv, d. h. einen Rückfall 'recidivate' hinter den (auf dem ursprünglichen Leistungsniveau) erreichten Gesundheitszustand erleiden. Oder ob sich das geheilte Gewebe in der alten Form und an derselben Stelle wieder verletzt 'reinjure, re-injure, rate of recurrent incidence', d. h. die alte Verletzung wieder auftritt 'recurrence'. Die Prüfzeit dient dazu den Behandlungserfolg im Sinne des Erfolgsmerkmals auf ein Rezidiv zu kontrollieren. Mithilfe des Trunkierungszeichens \* werden (lange) Wortketten aus ähnlichen Begriffen mit gleichem Wortstamm vermieden. Mit (return to) performanc\* kann die erfolgreiche Rückkehr bei Rennpferden zu dem früheren Leistungsniveau beschrieben werden; es umfasst die beiden Varianten performanc(s). Durch die Beschränkung auf den Suchbegriff 'suspensory' lässt sich die anatomischen Bezeichnungen 'suspensory ligament' eindeutig von 'accessory ligament (of ...flexor)' unterscheiden und letztere ausschließen. Die anatomischen Bezeichnungen 'teno-cyte(s), teno-lysis, teno-de-sis, teno-synovitis, teno-synov, teno-tomy' werden mit teno\* erfasst; mit tendo\* wird die Wortkette 'tendon(s), tendo-cyte(s), tendo-lysis, tendo-nitis, tendo-synovectomy, tendo-synovitis, tendo-vaginitis' erfasst; mit dem Suchbegriff tendi\* die Wortkette 'tendi-nitis, tendi-nocyte, tendi-nopathy, tendi-nous'. Unter Vermeidung bibliographischer Dubletten ergibt die formelle Suchanfrage am 09.01.2016 mit dem Filter 'Title/Abstract' 53 Ergebnisse. Der Suchstring lautet: *((((hors\*[Title/Abstract] OR equine\*[Title/Abstract])) AND (tendo\*[Title/Abstract] OR tendi\*[Title/Abstract] OR teno\*[Title/Abstract] OR flexor[Title/Abstract] OR suspensory[Title/Abstract])) AND (return[Title/Abstract] OR performanc\* [Title/Abstract])) AND ("2011"[Date - Publication]: "2016"[Date - Publication])*. Dieselbe Suchanfrage am 8.02.17 ergibt 54 relevante Ergebnisse, also 1 Studie mehr, nämlich GEBUREK et al. (2016). Die Suchanfrage erfolgt bis zum Abschluss der Arbeit jeweils zum Monatsbeginn automatisch.

#### 4.1.5 Suchstrategien in der Datenbank CAB Abstracts

Die Schlagwortsuche in CAB gestaltet sich ähnlich wie bei der PubMed-Suche. Der Organism-Descriptor 'od: horses' wird verknüpft mit nachgeordneten zielführenden Suchbegriffen. Der Suchstring: *((tendon\*) OR (suspensory\*)) AND (((therap\*))AND((od:(„horses“)) AND (((prp) OR („platelet rich plasma“) OR („stemcell\*))yr:[2011 TO 2016]* ist auf dem Screenshot ersichtlich und ergibt 59 Treffer.



Abb. 4.1.5 Suchstring in der Datenbank CAB Abstracts

The screenshot shows the CAB Direct search interface. At the top, there's a search bar with the text: `((tendon*) OR (suspensory*)) AND (((therap*)) AND ((od:("horses"))) AND (((prp) OR ("platelet riched plasma")) OR ("stem cell*"))yr:[2011 TO 2016]`. Below the search bar, there are tabs for "Search results", "My records", "My searches", "My CABI", and "My Projects". The "Search results" tab is active, showing "Search Results: 59 results (approx.)". There are filters for "Refined by: Language: English", "OR German", "OR Spanish", and "OR Italian". A "Results by Year" section on the right shows a slider from 2011 to 2016. The first search result is: "1. Tenogenically induced allogeneic mesenchymal stem cells for the treatment of proximal suspensory ligament desmitis in a horse."

#### 4.1.6 Suche nach thematisch verwandten Metaanalysen in der Pferdemedizin

Um abzuklären, ob thematisch verwandte Metaanalysen bereits vorliegen, wird bei PubMed jährlich am 22. November von 2015 bis 2017 mithilfe des Suchstrings: *"Meta-Analysis" [Publication Type] AND "Horses"[Mesh]* die Suchanfrage durchgeführt. Es ergeben sich 18 Treffer. Nur 3 davon sind Metaanalysen, jedoch mit anderer Thematik (SULLIVAN et al. 2015; MUNSTERMAN et al. 2016; CALZETTA et al. 2017). Da ARLT (2013) noch keine Metaanalyse zur Thematik Pferd findet, scheint SULLIVAN et al. (2015) die erste ihrer Art zu sein.

## 4.2 Arbeitsmethoden

Die Arbeitsfrage wird anhand von 30 Studien im Zeitraum 2011-2016 untersucht; zusätzlich anhand der überarbeiteten Ergebnisse von 21 Studien (1983-2011) aus der metaanalytischen Untersuchung von GEBUREK und STADLER (2011b). 12 (der 30) sind nicht randomisierte klinische Studien mit einer Vergleichsgruppe als Kontrollgruppe, 15 (der 30) sind unkontrollierte nicht randomisierte klinische Studien, 3 (der 30) sind kleine RCTs. Die Arbeit gliedert sich methodisch in zwei Teile, die sich notwendig ergänzen. So wird beispielsweise im I. Teil aus den extrahierten Primärdaten die 'Gebrauchsfähigkeitsrate' *GFR* berechnet, diese aber erst im II. Teil definiert.

### 4.2.1 Arbeitsschritte in Teil I: Extraktion der BehG - qualitative und quantitative Analyse

Teil I: Die 'gepoolte Re-Analyse' der Studien 2011-2016 gliedert sich in 3 Arbeitsschritte. (1) Definierte Einflussfaktoren und Merkmale der GF der Sportpferde werden aus den Studien extrahiert und in BehG geordnet. *GFR* wird berechnet. Dies sind die Primärdaten, sie werden tabellarisch präsentiert und deskriptiv analysiert. (2) Die BehG werden anhand qualitativer und quantitativer Vergleichskriterien auf Poolbarkeit untersucht. Die *GFR* poolbarer BehG werden mittels quantitativer Heterogenitätsanalyse gepoolt und der resultierende Effektschätzer *GFR<sub>gepoolt</sub>* (incl. Konfidenzintervall) ermittelt. (3) Unterschiede zwischen den Behandlungsformen (KtrB vs. PrTB; PrTB<sub>1</sub> vs. PrTB<sub>2</sub>) werden anhand der Effektdifferenz *d<sub>GFR</sub>* zwischen 2 vergleichbaren BehG in sog. paarweisen Vergleichen (**pV**) mittels Power-Analyse (Power 80 %;  $\alpha = 0,05$ ) per Signifikanznachweis statistisch gesichert, falls die erforderliche Fallzahl gegeben ist.

#### 4.2.2 Arbeitsschritte in Teil II: Methodisch systematische Entwicklung des *GFR*-Modells

Teil II: 'Das Evaluationsmodell *GFR*' wird systematisch methodisch in 3 Arbeitsschritten entwickelt. Da das Modell evidenzbasierte Entscheidungen ermöglichen soll, ist die Vergleichbarkeit der Ergebnisse für die Effektgröße *GFR* durch die interne Systematik der quantitativen Messung zu gewährleisten. Dazu wird die Berechnung von *GFR* normiert (5.5), die semiquantitativen und quantitativen diagnostischen Parameter bewertet (5.6) und die Messbedingungen für *GFR* kontrolliert, indem die Studiendauer differenziert nach Gewebestruktur OBS, TBS einerseits und FT andererseits standardisiert wird (5.7).

#### 4.2.3 Statistische Methoden

##### 4.2.3.1 Computerprogramme

Die erforderlichen statistischen Berechnungen (fixed-effects Modell, random-effects Modell, 95 %-Vertrauensintervalle, resultierende Effektgrößen  $GFR_{\text{gepoolt}}$ , Cochran's Q,  $I^2$ ; ebenso die Test-Verfahren F, t,  $\chi^2$ , z) werden mit der Software MedCalc durchgeführt (Version: update 1. Mai 2017; Kontakt: Acacialaan 22, 8400 Ostend, Belgium; MedCalc Software's VAT registration number is BE 0809 344 640). Die Power-Analyse (Fisher's exact test) wird mit dem freien Programm G\*Power 3(2007) durchgeführt (z. B. Website der Heinrich-Heine-Universität Düsseldorf).

##### 4.2.3.2 Statistische Heterogenitätsanalyse

In Teil 4.2.3.2-5 werden die Grundlagen der statistischen Heterogenitätsanalyse und danach in 4.2.3.6 die der Power-Analyse für das *GFR*-Modell beschrieben.

Wie kann der wahre Behandlungseffekt *GFR* mehrerer poolbarer BehG präzise geschätzt werden und wie sicher ist der ermittelte Schätzwert? Mittels statistischer Heterogenitätsanalyse wird ermittelt, in welchem Maß die Effekte  $GFR_i$  ( $i = 1, \dots, k$ ) von  $k$  inhaltlich homogenen, auf Bias geprüften BehG ausreichend quantitativ ähnlich sind um sie zu einem resultierenden *GFR*-Effektschätzer zusammenfassen zu können. D. h. ob und in welchem Umfang allein zufällige Abweichungen für die Unterschiede zwischen ihnen verantwortlich sind. Der Signifikanztest auf Homogenität der  $GFR_i$ -Effektgrößen sollte positiv ausfallen, wenn in einer Stichprobe aus mehreren BehG neben der Intervention keine weiteren Einflussfaktoren wirken. Andernfalls sollten Heterogenitätsquellen per Sensitivitätsanalyse detektiert werden; es kommen theoretisch plausible oder empirisch belegte Einflussfaktoren in Frage. Dazu können die BehG in Subgruppen unterteilt und untersucht werden in welchem Maß sie dazu beitragen, dass der Gesamteffekt *GFR* der Stichprobe robust und die Varianz der resultierenden Effektgröße verringert ist. Heterogenitätsquellen können auch durch Verfahren der Meta-Regression aufgedeckt werden (PREUSS 2013). Gelingt es die Auswirkungen inhaltlich substanzwissenschaftlich zu begründen, stärkt dies die Aussagekraft und Glaubwürdigkeit des Ergebnisses der Metaanalyse (PETERS et al. 2006; ZIEGLER et al. 2007; LEAN et al. 2009). Möglicherweise werden neue Hypothesen generiert sowie weiterführende Forschungsfragen angestoßen (ZIEGLER et al. 2007; LEAN et al. 2009).

Bei einer formalen Bewertung der Heterogenität, sollte das Heterogenitätsmaß, um vergleichbar zu sein, nicht von der Anzahl der BehG abhängen. Es stehen mehrere statistische Modelle zur Wahl, die sich im Heterogenitätsmaß für die Variabilität der  $GFR_i$  unterscheiden: fixed-effects Modelle (FE), random-effects Modelle (RE) oder zwei Regressionsmodelle speziell für Metaanalysen. Bei letzteren beiden, dem Meta-Regressionmodell und gemischten Modell, handelt es sich um Erweiterungen jeweils des FE bzw. RE. Das gemischte Modell kommt dann zum Einsatz, wenn die Heterogenität durch die Beziehung bestimmter Charakteristika (Kovariaten) der BehG zur Größe des Gesamteffekts mit dem Modell der Meta-Regression nicht erklärt werden kann (PREUSS 2013).

Im Folgenden (4.2.3.3-4) werden das FE und RE für das  $GFR$ -Modell erläutert. Die Ausführungen stützen sich, wenn nicht andere Quellen zitiert werden, wesentlich auf die Arbeit von PREUSS (2013); das Summenzeichen  $\sum$  gilt für  $i = 1, \dots, k$ .

#### 4.2.3.3 Fixed-Effects Modell

Bei fixed-effects Modellen (FE) wird angenommen, dass alle in  $k$  BehG beobachteten Werte für die Variablen A, B, C (5.5) derselben Grundgesamtheit (Population) entstammen. Unterschiede zwischen den  $GFR_i$  beruhen ausschließlich auf zufälligen Variationen in den BehG und folgen einer Normalverteilung um den wahren Wert  $GFR$  der Effektgröße. Unterschiede zwischen den  $k$  BehG gehen also allein auf die Standardfehler (Stichprobenfehler) der  $k$  Effektgrößen  $GFR_i$  ( $i = 1, \dots, k$ ) zurück. Der Standardfehler  $SE(GFR_i)$  der  $i$ -ten BehG berechnet sich aus der Größe  $n$  der BehG und der empirischen Varianz  $\sigma_i^2$  gemäß  $SE(GFR_i) = (\sigma_i^2 / n)^{1/2}$ . Heterogenitätsquellen zwischen den BehG darüber hinaus bezieht das Modell nicht ein.

Beim FE steht die Frage im Mittelpunkt, welcher Schätzwert  $GFR_i$  dem wahren Behandlungseffekt  $GFR$  am nächsten kommt. Das 95 %-CI des aus mehreren BehG gepoolten Effektschätzers  $\hat{GFR}_{FE}$  beschreibt die Unsicherheit bei der Schätzung des wahren Behandlungseffekts  $GFR$ .

Der gepoolte Effektschätzer ist:  $\hat{GFR}_{FE} = \sum(w_i \times GFR_i) / \sum w_i$  mit  $w_i = 1/\sigma_i^2$ .  $\hat{GFR}_{FE}$  ergibt sich als Quotient aus der Summe aller gewichteten Effekte  $GFR_i$  und der Summe aller Gewichte  $w_i$  und ist somit das gewichtete Mittel (bzw. der gewichtete Durchschnitt) der Effektschätzer  $GFR_i$  der  $k$  BehG. Die Gewichtung  $w_i$  berechnet sich als Kehrwert der Varianz  $\sigma_i^2$  der  $GFR_i$  (Inverse-Varianz-Ansatz).  $[\hat{GFR}_{FE} - t \times SE(\hat{GFR}_{FE}) ; \hat{GFR}_{FE} + t \times SE(\hat{GFR}_{FE})]$  ist das 95 %-Konfidenzintervall mit dem Standardfehler  $SE(\hat{GFR}_{FE}) = 1 / (\sum w_i)^{1/2}$  und dem kritischen Wert der Standardnormalverteilung  $t = 1,96$ . Mithilfe des Standardfehlers  $SE(\hat{GFR}_{FE})$  lässt sich für  $\hat{GFR}_{FE}$  eine Teststatistik  $Z_{FE}$  (Prüfgröße) konstruieren:  $Z_{FE} = \hat{GFR}_{FE} / SE(\hat{GFR}_{FE})$ .  $Z_{FE}$  ist approximativ einer Normalverteilung. Unter der Nullhypothese  $H_0$  kann mit der Prüfgröße  $Z_{FE}$  die zufallsbedingte Heterogenität zwischen den BehG durch eine modifizierte  $\chi^2$ -Teststatistik nachgewiesen werden. Dazu lässt sich z. B. Cochran's Q-Test verwenden.

$Q = \sum w_i \times GFR_i^2 - (\sum w_i \times GFR_i)^2 / \sum w_i = \sum w_i \times (GFR_i - \hat{GFR}_{FE})^2$  ist die klassische Cochran's Q-Teststatistik für das FE Modell. Die Nullhypothese  $H_0$  lautet 'es gibt keine Heterogenität', d. h. die wahren Effekte aller BehG sind gleich. Die Alternativhypothese  $H_1$  lautet 'mindestens eine Effektgröße  $GFR_i$  unterscheidet sich von den anderen'.  $Q$  folgt unter  $H_0$  einer  $\chi^2$ -Verteilung mit  $k-1$  Freiheitsgraden, woraus der zugehörige p-Wert ermittelt werden kann. Ein signifikantes Q-Testergebnis (z. B.  $p < 0,1$ ; Signifikanzniveau  $\alpha = 0,1$ ) deutet auf nicht unerhebliche Heterogenität;  $H_0$  wird verworfen.

Der Q-Test hat folgende Schwächen (TIMMER und RÜCKER 2008): Er prüft anhand des p-Wertes nur, ob alle BehG einen gleich großen Behandlungseffekt haben, d. h. er weist die Existenz von Heterogenität nach, bestimmt jedoch nicht das Heterogenitätsmaß, d. h. „wie viel“ Heterogenität besteht. Aus einem nicht-signifikanten Testergebnis kann nicht auf Homogenität geschlossen werden. Außerdem kann bei einer geringen

Anzahl BehG  $k < 10$  die Teststärke schwach sein, so dass tatsächlich stark abweichende  $GFR_i$  nicht das Testergebnis 'heterogen' herbeiführen. Bei einem BehG-Paar (vgl. 5.4.2.2) mit  $k = 2$  ist der Q-Test somit wenig sensitiv und hat eine geringe Power. Da Heterogenität grundsätzlich mit der steigenden Anzahl  $k$  der BehG zunimmt, gilt es zu beachten, dass der Test bei einer großen Anzahl von BehG in seiner Mächtigkeit so sehr anwachsen kann, dass tatsächlich gering abweichende  $GFR_i$  das Testergebnis 'heterogen' fälschlicherweise herbeiführen. Aufgrund der starken Schwankungen der Mächtigkeit des Tests ist es schwierig ein passendes festes Signifikanzniveau zu wählen (PREUSS 2013).

#### 4.2.3.4 Random-Effects Modell

Das random-effects Modell (RE) ist eine Erweiterung des FE und das Standardmodell bei Metaanalysen im Falle von Heterogenität zwischen den BehG; die Heterogenität zwischen den BehG wird mit  $\tau^2$  quantifiziert (PREUSS 2013). Wenn das FE zur Analyse von Heterogenität eingesetzt wird, die das Maß der Zufallsvariabilität übersteigt, erzeugt dies (fälschlicherweise) zu kleine Konfidenzintervalle und einen Anstieg des empirischen  $\alpha$ -Fehlers (Fehler 1. Art). Anders als das FE, das die Variabilität der  $GFR_i$  mit dem Standardfehler der Stichproben erklärt, von denen angenommen wird, dass sie alle aus ein und derselben Grundgesamtheit gezogen werden, geht das RE davon aus, dass die  $GFR_i$  der BehG aus Stichproben berechnet werden, die aus einer Menge unterschiedlicher Grundgesamtheiten stammen. D. h. die  $GFR_i$  entstammen aus einer übergeordneten (Normal-)Verteilung mit dem wahren Behandlungseffekt  $GFR$  und der Varianz  $\tau^2$ . Die Heterogenität zwischen den BehG entsteht, da die wahren Effekte in jeder BehG nicht identisch sind.

Beim RE steht die Frage im Mittelpunkt, welcher durchschnittliche Gesamteffekt der  $GFR_i$  bei  $k$  BehG für den gepoolten Effektschätzer  $\hat{GFR}_{RE}$  erwartet werden kann und welches Maß an Glaubwürdigkeit sich in der Breite des Konfidenzintervalls von  $\hat{GFR}_{RE}$  niederschlägt. Das 95 %-CI ist  $[\hat{GFR}_{RE} - t \times SE(\hat{GFR}_{RE}); \hat{GFR}_{RE} + t \times SE(\hat{GFR}_{RE})]$  mit dem Standardfehler  $SE(\hat{GFR}_{RE}) = 1 / (\sum(w_i^{-1} + \tau^2))^{1/2}$  und dem kritischen Wert der Standardnormalverteilung  $t = 1,96$ . Im RE werden BehG mit größerer Schätzungenauigkeit stärker gewichtet. Gepoolte Effektschätzer haben dadurch breitere Konfidenzintervalle. Allerdings besteht im RE die Gefahr, dass  $GFR_i$  kleiner BehG gegenüber großen überbewertet werden, da die Gewichtung anhand der Größe der BehG entfällt. Dies droht insbesondere bei einem Paar aus einer sehr großen und einer sehr kleinen BehG (vgl. Kapitel 5.4.2.2). Anders als beim FE erscheinen dann beide Effekte  $GFR_1$  und  $GFR_2$  signifikant ähnlich, d. h.  $H_0$  wird beibehalten, obwohl die Unterschiede der Effekte tatsächlich groß sind (TIMMER und RÜCKER 2008).

Der gepoolte Effektschätzer gemäß Simonian und Laird ist:  $\hat{GFR}_{RE} = \sum(w_i^* \times GFR_i) / \sum w_i^*$  mit  $w_i^* = w_i + \tau^2 = 1 / (\sigma_i^2 + \tau^2)$ . Im Unterschied zum  $\hat{GFR}_{FE}$  wird beim  $\hat{GFR}_{RE}$  die Gewichtung  $w_i$  um die Heterogenitätsvarianz  $\tau^2$  erweitert. Die Varianz des Gesamteffektes entsteht sowohl durch die Zufallsvariation der Schätzung der  $GFR_i$  in den  $k$  BehG als auch durch die systematische Variation der  $GFR_i$  zwischen den  $k$  BehG. Die Heterogenität  $\tau^2$  zwischen den BehG wird unter Verwendung von Cochran's Q Statistik geschätzt und mit der inversen Varianz-Methode zu den für die BehG spezifischen Varianzen  $\sigma_i^2$  addiert. Die Teststatistik für  $\hat{GFR}_{RE}$  ist:  $Z_{RE} = \hat{GFR}_{RE} / SE(\hat{GFR}_{RE})$ . Sie ist approximativ einer Normalverteilung.  $\tau^2$  kann mit verschiedenen Methoden quantitativ geschätzt werden, u. a. mit der Maximum-Likelihood Methode oder der Restricted-Maximum-Likelihood Methode. Gewöhnlich wird die momentenbasierte Schätzung nach DerSimonian und Laird verwendet (PREUSS 2013):  $\tau^2 = \max \{ 0 ; [Q - (k-1)] / [\sum w_i - \sum w_i^2 / \sum w_i] \}$  mit  $Q = \sum w_i \times (GFR_i - \hat{GFR}_{FE})^2$ . Der Wertebereich ist  $0 \leq \tau^2$ .  $\tau^2$  hat den Wert Null, wenn die wahre Effektgröße  $GFR$  aller BehG gleich ist; das RE geht dann in das FE über. Heterogenität besteht bei  $0 < \tau^2$ .

Die Heterogenitätsvarianz  $\tau^2$  hat folgende Schwächen. Wie Cochran's Q hängt  $\tau^2$  von der Anzahl  $k$  sowie der Größe der BehG ab, wobei der Wert von  $\tau^2$  mit steigender Anzahl  $k$  der BehG nicht zunehmen muss (PREUSS 2013).  $\tau^2$  beschreibt das Ausmaß der Variabilität zwischen den beobachteten Effekten  $GFR_i$  der  $k$  BehG, allerdings nicht den Anteil. Dadurch scheidet es als brauchbares Heterogenitätsmaß für direkte Vergleiche zwischen verschiedenen Metaanalysen aus. Die Variabilität (Breite des 95 %-Konfidenzintervalls) von  $\tau^2$  kann sich bei einer niedrigen Anzahl  $k$  von BehG sehr erhöhen, da die Verteilung von  $\tau^2$  mithilfe der  $k$  vorhandenen BehG geschätzt werden muss. Dies erschwert die Interpretation von  $\tau^2$  (PREUSS 2013).

Dem wirken HIGGINS und THOMPSON (2002) mit dem Heterogenitätsmaß  $I^2$  entgegen. Wie eingangs bemerkt, sollte ein Heterogenitätsmaß um vergleichbar zu sein, unabhängig von der Anzahl der BehG sein. Außerdem sollte es möglichst präzise den Anteil der Variabilität zwischen den  $GFR_i$  beschreiben, der nicht durch Stichprobenfehler, also durch Zufall wie beim FE vorausgesetzt entstanden ist (PREUSS 2013).

Das Heterogenitätsmaß  $I^2 = \max \{0; 100 \% \times [Q - (k-1)] / Q \}$  mit  $Q = \sum w_i \times (GFR_i - \hat{GFR}_{FE})^2$  von HIGGINS und THOMPSON (2002) wird wie  $\tau^2$  aus Cochran's Q berechnet, ist aber von der Anzahl  $k$  der BehG unabhängig. Im Zähler des Terms steht der Betrag von  $Q$ , der den Freiheitsgrad  $(k-1)$  absolut übersteigt.  $I^2$  nimmt Werte zwischen 0 % und 100 % an und bestimmt das Maß an Inkonsistenz zwischen den Effekten  $GFR_i$  der  $k$  BehG bezüglich des daraus hergeleiteten Effektschätzers  $\hat{GFR}_{RE}$ .  $I^2$  ist der prozentuale Anteil der totalen Varianz, der auf die Heterogenität zwischen den BehG zurückgeht und auf systematischen Unterschieden, somit nicht auf Zufall beruht (PREUSS 2013).  $I^2 = 0$  gilt, falls der Wert von  $Q$  kleiner oder gleich dem Freiheitsgrad  $(k-1)$  ist. Die Variabilität der  $GFR_i$  ist dann vollständig auf zufällige Abweichungen zurückzuführen. Wie erwähnt ist dies der Fall, wenn die wahre Effektgröße  $GFR$  aller BehG gleich ist; das RE geht dann in das FE über.

Das  $I^2$  ist keine Teststatistik. Es sollte nicht als Kriterium verwendet werden um BehG ein- oder auszuschließen; HIGGINS (2008) weist ausdrücklich darauf hin, dass gerade BehG mit einem Effekt nahe des gepoolten Schätzers  $\hat{GFR}_{RE}$ , die ausgeschlossen werden,  $I^2$  reduzieren.

#### 4.2.3.5 Quantitative Kriterien der Heterogenitätsanalyse

Es gibt (noch) kein einheitliches Verfahren, wie statistische Heterogenität formal bewertet und die Ergebnisse quantitativ gemessen werden sollen.

Der Forest-Plot liefert grafische Anhaltspunkte. In der Mitte sind die Punktschätzer  $GFR_i$  in Form von Rechtecken dargestellt, wobei die Größe der Rechtecksfläche die Gewichtung der BehG repräsentiert. Die Schnurrbärte überdecken jeweils das 95 %-Konfidenzintervall. Der gepoolte Effektschätzer  $\hat{GFR}$  wird als Raute dargestellt, deren waagrecht angeordnete Diagonale das 95 %-Konfidenzintervall exakt überdeckt. Ist das Konfidenzintervall klein, erscheint die Raute in horizontaler Richtung gestaucht, andernfalls gestreckt. Der Forest-Plot zeigt übersichtlich, wie weit die  $GFR_i$  auseinander liegen und bei wie vielen davon die Konfidenzintervalle sich überschneiden. Der Beitrag jeder einzelnen  $GFR_i$  zum resultierenden Gesamteffekt  $\hat{GFR}$  wird so ersichtlich. Auch gemeinsame Trends zwischen den  $GFR_i$  treten durch gruppierte Anordnung grafisch deutlich hervor. Über das grafische Hilfsmittel hinaus gibt es quantitative Untersuchungsmethoden, z. B. Anpassungstests, um die Vergleichbarkeit der BehG zu prüfen. Wenn wie in der vorliegenden Metaanalyse häufig nur zwei BehG zu einem Paar gepoolt werden (5.4.2.2), ist der Forest-Plot wirkungslos und scheidet aus.

Wie TIMMER und RÜCKER (2008) darlegen, ist die mangelnde Güte von Heterogenitätsstatistiken ein grundlegendes Problem, zumal bei kleiner Anzahl  $k$  der BehG. Das Problem besteht wie erwähnt insbesondere bei der Analyse der Paare mit  $k = 2$  (vgl. 5.4.2.2).  $I^2$  ist zwar nicht abhängig von der Anzahl  $k$  der BehG, jedoch von der Behandlungsgruppengröße (TIMMER und RÜCKER 2008; ZIEGLER et al. 2007; PREUSS 2013). Dies ist im *GFR*-Modell die Anzahl der regelkonform behandelten Patienten  $N-A$ . Steigt diese an, werden (bei festem  $\tau^2$  in einer gegebenen Metaanalyse) die Konfidenzintervalle der  $GFR_i$  kleiner, die Heterogenität zwischen den  $GFR_i$  aber wird größer mit der Folge, dass auch der Wert  $I^2$  größer wird (PREUSS 2013; TIMMER und RÜCKER 2008). Für die Interpretation der Heterogenitätsskala von  $I^2$  zwischen 0 % und 100 % lassen sich keine exakten Grenzwerte definieren. Lediglich Skalen-Bereiche (d. h. offene Intervalle) können Stufen der Heterogenität tendenziell zugeordnet werden. HIGGINS und THOMPSON (2002) empfehlen:  $I^2 < 25$  % deutet auf niedrige Heterogenität;  $I^2$  ca. 50 % deutet auf moderate Heterogenität; 75 %  $< I^2$  deutet auf hohe Heterogenität. DEEKS, HIGGINS und ALTMANN (2011) empfehlen im Cochrane Handbuch, Kapitel 9, eine Zuordnung mit fließenden Bereichsgrenzen: 0-40 % weist auf möglicherweise unbedeutende Heterogenität; 30-60 % kann moderate, 50-90 % kann erhebliche Heterogenität repräsentieren; 75-100 % bedeutet beträchtliche Heterogenität.

Ersichtlich haben bei der Heterogenitätsanalyse in vielen Fällen, insbesondere wenn  $I^2$  Werte um 50 % annimmt, bei der Entscheidung für oder gegen die Poolbarkeit gute Sachkriterien Priorität vor rein rechnerisch 'richtigen' quantitativen Methoden. Daher werden in 5.4.2.1 Sachkriterien entwickelt. Wie erwähnt bleibt zu beachten, dass (a) der Q-Signifikanztest für eine kleine BehG-Anzahl  $k$  mit  $k < 10$  nur schwach sensitiv ist; (b) die Aussagekraft des p-Wertes u. a. mit der Anzahl  $k$  der BehG steigt; (c)  $I^2$  zwar nicht von der Anzahl  $k$ , jedoch von der Größe der Fallzahl  $N-A$  der BehG abhängt. In solchen Fällen wird deutlich, dass statistische formale Berechnungs- und Darstellungsverfahren substanzwissenschaftliche Überlegungen nicht ersetzen können. Hier müssen die Ursachen für Heterogenität, etwa klinische oder methodische Faktoren oder unterschiedliche Zusammensetzungen der BehG geklärt werden, damit die Entscheidung zwischen den Modellen transparent und sachlich nachvollziehbar ist.

#### 4.2.3.6 Power-Analyse der Effektdifferenz zwischen BehG

Es stellt sich die Frage, wie valide geprüft werden kann, ob die Effektdifferenz  $d_{GFR} = GFR_1 - GFR_2$  zwischen 2 BehG auf einen zufälligen oder systematischen Unterschied der therapeutischen Wirkung verschiedener Behandlungsformen bzw. prT – bei sonst gleichen Bedingungen – hindeutet. Ein '*GFR*-Goldstandard' als Vergleichswert liegt noch nicht vor. Wahrscheinlich eignet sich dafür die *GFR* der Behandlungsform KtrB; der dafür (5.4.3) ermittelte empirische Wert  $GFR = 57,5$  % [44,4 %-69,9 %],  $N-A = 60$  ist nur ein erster Näherungswert.

Die Power-Analyse (COHEN 1988) ermöglicht die Effektdifferenz  $d_{GFR}$  zwischen 2 BehG, sog. paarweise Vergleiche (pV), bei vorgegebener Power (z. B. 80 %) auf einem bestimmten Signifikanzniveau (z. B.  $\alpha = 0,05$ ) mit der erforderlichen Fallzahl per Signifikanznachweis statistisch zu sichern. Die Power oder statistische Mächtigkeit ist die Wahrscheinlichkeit den Unterschied der Effekte zu entdecken, falls er tatsächlich vorhanden ist. Gelingt der Nachweis, so deutet die Effektdifferenz auf eine wahrscheinlich unterschiedliche Erfolgs-Chance *OR* zwischen beiden Behandlungsformen. Gelingt der Nachweis nicht, bedarf es zur Klärung weiterer Studien mit der größeren erforderlichen Fallzahl.

Die Power ist definiert als  $1-\beta$ , wobei  $\beta$  (z. B. = 20 %) der Fehler 2. Art ist. COHEN (1988) fordert, dass der  $\beta$ -Fehler maximal das Vierfache des Signifikanzniveaus  $\alpha$  betragen sollte; ist  $\alpha = 0,05$  folgt daraus für die Power 80 % (=  $1-\beta$ ). Die Power 80 % erscheint COHEN (1988) ein brauchbares Minimum. Das bedeutet, dass durchschnittlich in 80 % aller Fälle, bei denen identische BehG mit der erforderlichen Fallzahlgröße untersucht werden, der Unterschied des Behandlungserfolgs zwischen ihnen, wenn er tatsächlich existiert, entdeckt und der Irrtum auf  $\beta = 20$  % begrenzt ist. Das gewählte Signifikanzniveau  $\alpha$  legt das maximale Risiko für den Fehler 1. Art fest. Er entsteht dadurch, dass die Nullhypothese, z. B. „ $GFR_1$  repräsentiert einen gleich großen Behandlungserfolg wie  $GFR_2$ “, zufällig abgelehnt wird, obwohl sie tatsächlich zutrifft.

Weichen die Effekte  $GFR_1$  und  $GFR_2$  zweier prT stark voneinander ab, ist eine relativ kleine Fallzahl erforderlich, d. h. wenige Patienten um bei vorgegebener Power (z. B. 80 %) die relativ große Effekt-Differenz statistisch zu sichern. Dagegen werden bei kleinen Effekt-Differenzen relativ große Stichproben dafür benötigt. Haben andererseits z. B. 2 BehG nur Patienten mit minimalen Läsionen (LG 1-3) mit sehr guter Heilungsprognose, so interessiert eine relativ große Effekt-Differenz, dagegen bei leichten bis mittleren Läsionen (LG 4-5) eine im Vergleich dazu kleinere. Bei schweren Läsionen LG 6 oder Rupturen mit ungünstiger Heilungsprognose fällt bereits eine kleine Effekt-Differenz klinisch und systematisch ins Gewicht.

#### 4.2.4 Reduzierung von Fehlerquellen und Bias-Risiken

##### 4.2.4.1 Typische Fehler in Metaanalysen

Einige wichtige Fehlerquellen bei der Planung, Durchführung und Auswertung von Studienergebnissen in Metaanalysen sollen bei der vorliegenden Metaanalyse vermieden werden.

Dazu zählt an prominenter Stelle das 'Äpfel-und-Birnen-Problem' (SACKETT et al. 1996), auch als Inkommensurabilitäts oder (vice versa) Uniformitätsproblem bezeichnet. Es beschreibt die Gefahr, dass sich die Studien in zentralen Merkmalen wie Studiendesign, Operationalisierung der abhängigen Variablen etc. unterscheiden. Bei der vorliegenden Metaanalyse entsteht das Problem der validen Messung des Behandlungserfolgs mit prT, weil Primärstudien nicht dasselbe Messverfahren oder dieselbe Definition der Effektgröße verwenden. Dieses Problem soll entschärft werden, indem BehG mit vergleichbaren definierten Einflussfaktoren extrahiert werden und die Berechnungsweise der Effektgröße  $GFR$  aus drei disjunkten Variablen (5.5) normiert wird. Außerdem werden gleiche Messbedingungen hergestellt, indem Studien- und Prüfzeitdauer standardisiert werden.

Eine weitere typische Fehlerquelle bemisst sich nach dem Kriterium methodologischer Heterogenität. Damit Metaanalysen valide Ergebnis erzielen, müssen sämtliche Studien mit oder ohne signifikanten Ergebnissen untersucht werden. Letztere verbleiben jedoch häufiger in der sog. „Schublade“. Die Fehlerquelle wird üblicherweise als „Schubladenproblem“ oder „file drawer“ bezeichnet (PETERS et al. 2006). Die selektive Studiauswahl erzeugt einen Publikationsbias. Mit dem statistischen Fail-Safe-N-Verfahren ist es möglich Art und Größenordnung des Problems rechnerisch näherungsweise darzustellen (ARLT 2013). In der vorliegenden Metaanalyse erscheint die Fehlerquelle vermeidbar, da aus den vorhandenen Studien sämtliche BehG ausgewertet werden, d. h. große und kleine mit unpräzisen  $GFR$ .

Eine weitere typische Fehlerquelle ist das als „Garbage in – garbage out“ bekannte Problem. Der Schwachpunkt ist, dass Studien unterschiedlicher methodischer Qualität miteinander in Beziehung gesetzt werden (ARLT 2013). Die „schlechten“ Studien darunter verzerren durch ihre Bias-Behaftung das Ergebnis der Metaanalyse. Dem Fehler wird vorgebeugt, indem die interne Qualität der BehG, bevor sie gepoolt werden, anhand der Studien, aus denen sie extrahiert werden, auf Bias geprüft und das Ausmaß der Bias-Behaftung bewertet wird (vgl. 4.2.4.4).

#### 4.2.4.2 Vier häufige Formen von Bias in den Studien

Bias in den untersuchten Studien und damit auch in den extrahierten BehG erzeugt eine Verzerrung der Ergebnisse der Metaanalyse, d. h. die systematische Abweichung vom wahren bzw. tatsächlichen Wert. Im Gegensatz zu zufälligen Fehlern, die sich mit zunehmender Größe der Stichprobe aufheben, addieren sich systematische Fehler. Speziell durch das Studiendesign der randomisierten kontrollierten Studie (RCT) soll Bias methodisch reduziert werden. Übersichtlich wird dies bei GREENHALG (2000) dargestellt. Er kategorisiert Bias nach vier Entstehungsursachen. (1) Verzerrung durch selektive Auswahl der Studienteilnehmer führt zu Selektionsbias; Lösungsansatz zur Reduktion: Randomisierung der Patienten. (2) Mängel bei der Zuteilung der Patienten oder Behandler zur Interventions- und Kontrollgruppe führen zu Performance Bias; Lösungsansatz zur Reduktion: Verblindung der Gruppen. (3) Fehler durch subjektive Bewertung der Messdaten führen zu Detection Bias; Lösungsansatz zur Reduktion: Verblindung der Bewerter. (4) Verlust von Studienteilnehmern während des Studienverlaufs, so dass die restliche Studienpopulation durch ein bestimmtes Merkmal systematisch geprägt ist, führt zu Attrition Bias; Lösungsansatz zur Reduktion: Strikte Einhaltung des Studienprotokolls. Jede Form von Bias beeinflusst die interne Validität des Untersuchungsergebnisses. Daher sollten Studien, deren Ergebnisse in Metaanalysen integriert werden, auf ihre Anfälligkeit für systematische Fehler geprüft werden (SANDERSON et al. 2007). Das gilt entsprechend als Grundsatz auch für die extrahierten BehG.

#### 4.2.4.3 Randomisierte kontrollierte klinische Studien – RCTs

Wie erwähnt sind unter den ca. 51 Studien seit 1983 nur 3 kleine RCTs, die übrigen haben eine Vergleichsgruppe als Kontrollgruppe oder sind in der Mehrzahl unkontrollierte Studien. SCHULZ und GRIMES (2005) stellen fest, dass solide durchgeführte RCTs selbst bei kleiner Fallzahl, geringer Power und geringer Präzision im Wesentlichen unverzerrte Effektschätzer liefern; RCTs mit geringer Power sollten daher unabhängig von ihren Ergebnissen publiziert werden. Die Überlegungen lassen sich – so SCHULZ und GRIMES (2005) – mit dem Konzept prospektiv geplanter Metaanalysen verbinden. Dies bedeutet für das *GFR*-Modell, dass aus mehreren RCTs homogene Bias freie kleine BehG mit *GFR* und Intervallschätzer extrahiert werden können und daraus wiederum ein resultierender Effektschätzer gepoolt werden kann, anhand dessen sich der Behandlungserfolg quantitativ valide bewerten lässt.

#### 4.2.4.4 Nicht randomisierte kontrollierte bzw. nicht randomisierte nicht kontrollierte klinische Studien

Wie lässt sich das Bias-Risiko im *GFR*-Modell auch bei klinischen Interventionsstudien (1) ohne Randomisierung und mit einer Vergleichsgruppe als Kontrollgruppe oder (2) ohne Randomisierung und ohne Vergleichsgruppe kontrollieren? Von den 30 Studien 2011-2016 sind 12 nicht randomisiert mit einer Vergleichsgruppe als Kontrollgruppe und 15 unkontrolliert und nicht randomisiert. Das ernsthafte Problem des Bias-Risiko erscheint im zweiten Fall verschärft. Aus der Perspektive des *GFR*-Modells rekrutieren sich die Patien-



ten aller untersuchten Studien idealerweise aus einer nicht vorselektierten Grundgesamtheit, aus der bei der metaanalytischen Untersuchung die extrahierten BehG als Stichproben 'zufällig' gezogen werden. Bekanntlich ist nicht eine kleine Fallzahl das Hauptproblem klinischer Studien, sondern die Verzerrung der Stichprobe durch systematische Fehler (SCHULZ und GRIMES 2005; MATTEWS 1995). Werden inhaltlich homogene, auf Bias geprüfte BehG gepoolt, kann das Biasrisiko anhand der Studien, aus denen sie extrahiert werden, graduell bewertet, möglicherweise durch geeignete Subgruppenbildung begrenzt und der resultierende Effektschätzer  $GFR_{\text{gepoolt}}$  einschließlich Variabilität (Konfidenzintervall) mit steigender Anzahl der BehG genauer geschätzt werden.  $GFR$  von BehG mit einer großen Fallzahl und geringer Variabilität, d. h. kleinen Vertrauensintervallen können quantitativ stärker gewichtet werden (LEAN et al. 2009; ZIEGLER et al. 2007). Die erzielbaren Ergebnisse sind replizierbar und bieten die Chance auf hohe Evidenz. Dieser Weg wird, soweit sich poolbare inhaltlich homogene BehG ergeben, in der vorliegenden metaanalytischen Studie beschritten. Somit kann anhand von  $GFR_{\text{gepoolt}}$  (einschließlich Variabilität) die tatsächliche Wirkung eines prT zwar nicht wie bei RCTs auf höchster Evidenzstufe, jedoch genauer geschätzt werden als bisher mit einzelnen Studien, von denen jede ihr Ergebnis individuell bewertet. Die Vorgehensweise erinnert an die Diskussion, die seit Jahren in der Humanmedizin geführt wird, auf die ARLT (2013) jüngst wieder hinweist: Werden zur gleichen Fragestellung Ergebnisse von Metaanalysen von RCTs mit Resultaten von Beobachtungsstudien, Fallsammlungen oder Kohorten von Risiko- bzw. Behandlungsgruppen, verglichen so weicht das durchschnittliche Ergebnis der Beobachtungsstudien kaum von dem der Metaanalysen ab.

**Tabelle 5.1a** Vergleich von Studien (2011-2016) zur Behandlung natürlich entstandener Erkrankungen des **Sehnenapparates** OBS, TBS, USB der TBS entweder mit kontrollierter Bewegung allein oder kombiniert mit potenziell regenerativen Therapeutika

Behandelte Gewebe- struktur	Substrat	Studie			n Studien- popula- tion	N Anzahl behandelte Pferde
		Laufende Nr. Erstautor		Design		
<b>OBS</b>	Nur kontrollierte Bewegung (orthopädischer Hufbeschlag)	1. Coudry	2014	<i>S-Vgl</i>	51	16
	**	2. Tipton	2013	<i>S</i>	19	10
	**	3. Geburek	2015	<i>RCT</i>	15	5
	**	4. Van Loon <sup>+</sup>	2014	<i>S</i>	52	1
	PRP	5. Zuffova	2013	<i>S</i>	22	22
	**	6. Geburek	2016	<i>RCT</i>	20	10
	**	7. Abellonet <sup>+</sup>	2009	<i>S-Vgl</i>	191	44
	**	8. Scala <sup>+</sup>	2014	<i>S-Vgl</i>	150	45
	Konditioniertes Serum	3. Geburek	2015	–	–	8
	Blut-MSC	9. Marfe	2012	<i>S-Vgl</i>	3	3
	Natives Knochenmarkaspirat	10. Russel	2016	<i>S</i>	105	16
	Knochenmark-MSC	11. Lange-C. <sup>+</sup>	2013a	<i>S-Vgl</i>	95	32
	**	12. Iacono	2015	<i>S-Vgl</i>	10	5
	Knochenmark-MSC kombiniert mit PRP	13. Renzi <sup>+</sup>	2013	<i>S-Vgl</i>	33	9
	Fettgewebe-MSC	12. Iacono	2015	–	–	5
	Fettgewebe-MSC kombiniert mit PRP	14. Guercio	2015	<i>S</i>	9	9
	Konditioniertes Medium aus allog. Amniongewebe-MSC	15. Lange-C <sup>+</sup>	2013b	<i>S</i>	13	5
	**	16. Tassan <sup>+</sup>	2014	<i>S</i>	16	5
	Allogene Amniongewebe-MSC	11. Lange-C.	2013a	–	–	34
	Allogene Fettgewebe-MSC kombiniert mit autologem PRP	17. Ricco	2013	<i>S</i>	19	19
	Allogene Nabelschnurblut-MSC	4. Van Loon	2014	–	–	23
	**	18. Kang	2013	<i>S</i>	6	6
	<b>Xenogene</b> ovine Amnion-Epithel-Gewebe-MSC	19. Muttini	2013	<i>S</i>	15	12
	**	20. Muttini	2015	<i>S</i>	6	6
<b>TBS</b>	Nur kontrollierte Bewegung und orthopäd. Hufbeschlag	21. Lutter	2015	<i>S-Vgl</i>	118	18
	PRP	7. Abellonet	2009	–	–	6
	**	8. Scala	2014	–	–	8
	Knochenmark-MSC kombiniert mit PRP	13. Renzi	2013	–	–	5
	Knochenmark-MSC	11. Lange-C.	2013a	–	–	2
	Allogene Amniongewebe-MSC	11. Lange-C.	2013a	–	–	3
<b>USB</b> der TBS	Allogene Nabelschnurblut-MSC	4. Van Loon	2014	–	–	6
	PRP	8. Scala	2014	–	–	5
	Allogenes PRP	22. Beerts	2013	<i>S-Vgl</i>	8	4
	Sehnenzell-induzierte MSC aus allogenem Blut	22. Beerts	2013	–	–	4

**MSC** multipotent mesenchymal stromal cells (ISCT Internat. Society for Cellular Therapy, 2006); **OBS** oberflächliche Beugesehne; **PRP** platelet rich plasma; **RCT** randomisierte kontrollierte klinische Studie. **S** unkontrollierte klin. Studie; **S-Vgl** Studie mit Vergleichsgruppe als Kontrollgruppe; **TBS** tiefe Beugesehne; **UBM** porcine urinary bladder matrix; **USB** Unterstützungsband der tiefen Beugesehne.

**Sonstige Abkürzungen:** – dieselbe Angabe; \*\* dasselbe pT; + Studie behandelt auch OBS oder TBS oder USB der TBS. **LG** Läsionsgrad; **M** Monat(e); **Med** Median; **Mw** Mittelwert; **Pz** Prüfzeit; **Rd** Rehabilitationsdauer.

Tabelle 5.1a Fortsetzung

Nutzungsrichtung	Prüfzeit Pz	Erfolgsmerkmal Rehabilitationsdauer Rd Erfolgsrate			R Rezidivrate	GFR Gebrauchsfähig- keitsrate
	M	Durchschnitt /Spanne M				
1. Trab16	24	Rückkehr zu Rennen	9 Mw	87,5 %	k. A.	k. A.
2. Cutting Horses10	6-36	ur/höN	4 Mw/ 3-12	90 % 9/10	33 % 3/9	67 % 6/9
3. FL1 S1 E1 V2	24-48	ur/höN	6,75 -24	100 % 4/4	0 %	100 % 4/4
4. D/S 1	6 < Pz	ur/höN für mehr als 6 M	Rd < 6	100 %	k. A.	k. A.
5. FL22 11+9+2 mit LG 6-5-4	k. A.	Teilnahme an 1 < Rennen	11 Med/ 2-23	93 % 13/14 LG 6-5-4 64-56-100%	k. A.	k. A.
6. D5 R4 E1	24	ur/höN	6	5/7	1/5	57 % 4/7
7. Trab35 S7 R2	12-42	urN	3,5-5,5	80% 35/44 Trab25	20% 7/35 Trab6	64 % 28/44
8. Spo45	12 < Pz	urN in 6 M	Rd < 6	k. A.	k. A.	k.A
FL4 D1 V3	–	–	6,75 -24	63 % 5/8	0 %	63 % 5/8
9. D1 S2	36	urN	4-5	100 % 3/3	0 %	100 % 3/3
10. Th /St16	24	Teilnahme an 1 < Rennen	18	69 % 11/16	k. A.	k. A.
11. D8 S9 FL7 E5 Trab3	24	urN	4-12	84-100 % 27/32 -1	22-26% 7/32-7/27	63-78% 20/32-25/32
12. FL5	k. A.	urN	12	100 % 5/5	k. A.	k. A.
13. HR9	24	Rückkehr zu Rennen	8-15	100 % 9/9	33 % 3/9	67 % 6/9
FL5	–	–	–	100 %	k. A.	k. A.
14. S5 Ba3 Rg1	k. A.	Rückkehr zu Rennen	7-9	100 % 7/7	0 %	100 % 7/7
15. D2 S3	24	urN	4-15	100 %	1/5	80 % 4/5
16. S3 D2	12	urN	6-10	100 %	1/5	80 % 4/5
D9 S13 FL6 E4 Trab2	–	–	4-5	100 % 34/34	6 % 2/34	94 % 32/34
17. R16 FL2 V1	24	urN	12	90 % 17/19	k. A.	k. A.
D/S 23	–	–	–	87 % 20/23	k. A.	k. A.
18. FL6	k. A.	Ausfallrate* senken bz. LG 5	3-6	33 %	k. A.	k. A.
D1 E3 FL5 S3	6	ur/höN	12	100 %	17 % 2/12	83 % 10/12
20. FL4 D1 E1	6	Histol. Evaluation	6	100 % 6/6	k. A.	k. A.
21. R18	42 Med 4-72	urN	6	k. A.	k. A.	k. A.
Trab4 R2	–	–	–	100 %	17 % 1/6 Trab1	83 % 5/6
Spo8	–	–	–	k.A	–	k.A
HR5	–	–	–	100 % 5/5	40 % 2/5	60 % 3/5
S1 FL1	–	–	(4-12)	50-100 % ½ -1	0 %	50-100 % ½ -1
S2 FL1	–	–	(4-5)	100 % 3/3	0 %	100 % 3/3
D/S 6	–	–	–	67 % 4/6	k. A.	k. A.
Spo5	–	–	–	k.A	k.A	k.A
21. V4	k. A.	wieder Training	4	0 %	k. A.	k. A.
V4	–	–	–	75 %	k. A.	k. A.

**Nutzungsrichtungen:** **Ba** Barrel racing; **D** Dressur; **E** Eventing; **End** Endurance; **FL** Flachrennen; **HR** Hindernis-Rennen; **Rg** Reining; **R** Reiten (auch Freizeit); **S** Springen; **Spo** Sport in athletischen Disziplinen; **Th** Thoroughbred; **Trab** Trabrennen (Standardbred); **V** Verschiedene; **Vo** Voltegiere.

**ur/höN** Nutzung auf dem ursprünglichen bzw. höheren Leistungs-Niveau nach Rehabilitation; **x < y** „x weniger als y“ bzw. „y größer als x“.

\* Ausfallrate von Verletzungen der Kategorie LG 5 (LG nach GENOVESE et al. 1997) beträgt 90 %.

**LG** Läsionsgrad nach RANTANEN et al. (2011). - **Weitere Abkürzungen** vgl. Tabelle 5.1b.

Tabelle 5.1b Vergleich von Studien (2011-2016) zur Behandlung natürlich entstandener Erkrankungen des <b>Fesselträgers</b> FT entweder mit kontrollierter Bewegung allein oder kombiniert mit potenziell regenerativen Therapeutika						
Behandelte Gewebestruktur	Substrat	Studie			n	N
		Laufende Nr.		Design	Studienpopulation	Anzahl der behandelten Pferde
		Erstautor				
<b>FT</b>  Ursprung	PRP**	23. Romagnoli 24. Lutz	2015 2011	<i>S</i> <i>S-Vgl</i>	20 109	20 14
	Knochenmark-MSC kombiniert mit PRP	13. Renzi <sup>+</sup>	2013	<i>S-Vgl</i>	(33)	2
	Fettgewebe-MS	25. Rich	2014	<i>S</i>	83	30
	Sehn.-ind. MSC aus allogenen Blut und allogenes PRP	26. Vandenberghe	2015	<i>S</i>	1	1
	Porcine UBM (urinary bladder matrix)	24. Lutz	2011	—	—	14
<b>FT</b>  Körper	PRP	24. Lutz	2011	—	—	15
	Natives Knochenmarkaspirat	27. Hall	2013	<i>S</i>	30	9
	Fettgewebe-MSC	25. Rich	2014	—	—	6
	Knochenmark-MSC kombiniert mit PRP	13. Renzi	2013	—	—	2
	Konditioniertes Medium aus allog. Amniongewebe-MSC**	15. Lange-C. <sup>+</sup> 16. Tassan <sup>+</sup>	2013b 2013	<i>S</i> <i>S</i>	(13) (16)	2 2
	Porcine UBM	24. Lutz	2011	—	—	8
<b>FT</b>  Schenkel	Nur kontrollierte Bewegung	4. Van Loon <sup>+</sup>	2014	<i>S</i>	(52)	2
	Nur kontrollierte Bewegung und orthopäd. Hufbeschlag	28. Plevin	2014	<i>S-Vgl</i>	892	85
	PRP	29. Garrett	2013	<i>RCT</i>	39	20
	** (lateral)	24. Lutz	2011	—	—	20
	** (medial)	24. Lutz	2011	—	—	22
	Natives Knochenmarkaspirat	27. Hall	2013	—	—	20
	Fettgewebe-MSC	25. Rich	2014	—	—	15
	Knochenmark-MSC kombiniert mit PRP	13. Renzi	2013	—	—	2
	Konditioniertes Medium aus allog. Amniongewebe-MSC**	15. Lange-C. <sup>+</sup> 16. Tassan <sup>+</sup>	2013b 2013	— —	— —	5 6
	Allogene Nabelschnurblut-MSC	4. Van Loon	2014	—	—	22
	Porcine UBM (lateral)	24. Lutz	2011	—	—	3
	(medial)	24. Lutz	2011	—	—	7
<b>FT ohne</b>  Angabe der Lokalisation	PRP**	30. Banse	2011	<i>S-Vgl</i>	20	10
	**	7. Abellanet <sup>+</sup>	2009	<i>S-Vgl</i>	(191)	19
	**	24. Lutz	2011	—	—	71
	**	8. Scala <sup>+</sup>	2014	<i>S-Vgl</i>	(150)	35
	Fettgewebe-MSC	25. Rich	2014	—	—	32
	Knochenmark-MSC	11. Lange-C. <sup>+</sup>	2013a	<i>S-Vgl</i>	(95)	10
	Allogene Knochenmark-MSC	30. Banse	2011	—	—	10
	Allogene Amniongewebe-MSC	11. Lange-C.	2013a	—	—	14
Porcine UBM	24. Lutz	2011	—	—	34	

**FT** Fesselträger; **LG** Läsionsgrad nach RANTANEN et al. (2011); **PRP** platelet rich plasma; **RCT** randomisierte kontrollierte klinische Studie; **S** unkontrollierte klinische Studie; **S-Vgl** klinische Studie mit Vergleichsgruppe als Kontrollgruppe; **UBM** urinary bladder matrix.

**Sonstige Abkürzungen:** – dieselbe Angabe; \*\* dasselbe prT; + Studie behandelt auch OBS oder TBS oder USB der TBS; **h** hinten; **J** Jahr(e); **M** Monat(e); **Med** Median; **Mw** Mittelwert; **Pz** Prüfzeit; **Rd** Rehabilitationsdauer; **sehn.-ind.** sehnenzell-induziert.

Tabelle 5.1b Fortsetzung

Nutzungsrichtung	Prüfzeit Pz M	Erfolgsmerkmal		<i>E</i> Rehabilitationsdauer Erfolgsrate		<i>R</i> Rezidivrate	<i>GFR</i> Gebrauchsfähigkeitsrate
		M Durchschnitt /Spanne					
23. Spo20	6-24	urN	3,6 Med / 3-6	100 %		20 % 4/20	80 % 16/20
24. D4 S5 Trab1 V4	17 < P	urN	2-12	92 % 12/13		42 % 5/12	54 % 7/13
HR2	24	(Rückkehr zu Rennen)	(8-15)	100 %		0 %	100 % 2/2
25. Spo30	13-92	nach 5 M kontroll. Bewegung ur/höN und Rezidiv frei für 12 < M	5	87 % 26/30		0 %	87 % 26/30
26. D1	12	urN	8	100 %		0 %	100 % 1/1
D6 S3 Trab1 V4	—	—	—	69 % 9/13		44 % 4/9	38 % 5/13
D7 S3 Trab2 V3	—	—	—	100 % 15/15		20 % 3/15	80 % 12/15
27. G9	24-60	urN, 1 < Start	6 < Rd	67 % 6/9		k. A.	k. A.
Spo6	—	—	—	67 % 4/6		0 %	67 % 4/6
HR2	—	—	—	50 % ½		0 %	50 % 1/2
End2	24	(urN)	(4-15)	100 %		0 %	100 % 2/2
E2	12	(urN, Training)	12	100 %		0 %	100 % 2/2
D1 S1 Trab5 V1	—	—	—	100 %		13 % 1/8	88 % 7/8
D u. S 2	6 < Pz	(ur/höN)	(Rd < 6)	50 % ½		k. A.	k. A.
28. Th85 Jährlinge	10-18	Starts bis Ende der 3-jährigen Renn-Saison	6 < Rd	62 % 53/85		k. A.	k. A.
29. Th20 Jährlinge	6-48	Starts bis Ende der 4-jährigen Renn-Saison	6 M im Jährlings-Jahr	80 % 16/20		ca.10; 29; 38% der Fälle mit LG 4; 5; 6	ca.70; 56% der Fälle mit LG 4; 5
D11 S6 Trab1 V2	—	—	—	90 % 18/20		k. A.	k. A.
D6 S5 Trab5 V6	—	—	—	82 % 18/22		28 % 5/18	65 % 13/20
FL20	—	—	—	90 % 18/20		22 % 4/18	64 % 14/22
Spo15	—	—	—	79 % 11/14		k. A.	k. A.
HR2	—	—	—	100 %		0 %	79 % 11/14
D2 End1 Vo1 S1	24	(urN)	(4-12)	4/5		50 % 1/2	50 % 1/2
D2 S2 E1 V1	—	(urN)	4	5/5		1/4	60 % 3/5
D u. S 22	—	—	—	68 % 15/22		1/5	80 % 4/5
D3	—	—	—	100 %		k. A.	k. A.
D1 S1 Trab2 V3	—	—	—	86 % 6/7		33 % 1/3	67 % 2/3
30. V10	12	urN	6	44 %		0 %	94 % 29/31
Trab9 D2 S5 R3	(12-42)	(urN)	(4,5-6)	100 %		20-40% 1/5-2/5	30-80% 3/10-8/10
D28 S19 Trab9 V15	—	—	—	89 % 62/70			
Spo35	12 < Pz	(urN in 6 M)	(Rd < 6)	k. A.			
Spo32	—	—	—	94 % 29/31			
D4 S4 Trab2	24	(urN)	(4-12)	50-100% 5/10 -1			
V10	—	—	—	56 %		k. A.	k. A.
D5 S8 Trab1	—	—	(4-5)	100 %		0 %	100 % 14/14
D11 S5 Trab8 V10	—	—	—	84 % 27/32		26 % 7/27	63 % 20/32

**Nutzungsrichtungen:** **Ba** Barrel racing; **D** Dressur; **E** Eventing; **End** Endurance; **FL** Flachrennen; **HR** Hindernis-Rennen; **Rg** Reining; **R** Reiten (auch Freizeit); **S** Springen; **Spo** Sport in athletischen Disziplinen; **Th** Thoroughbred; **Trab** Trabrennen (Standardbred); **V** Verschiedene; **Vo** Voltegieren.

**ur/höN** Nutzung auf dem ursprünglichen bzw. höheren Leistungs-Niveau nach Rehabilitation; **x < y** „x weniger als y“ bzw. „y größer als x“.

Teil I gliedert sich grob wie folgt: (5.1) werden alle aus den Studien 2011-2016 extrahierten BehG präsentiert (Tabelle 5.1.a und 5.1.b; aus drucktechnischen Gründen vor Kap. 5.1 gestellt). Es folgen statistisch deskriptive Vergleiche zwischen den Studien im zeitlichen Verlauf 1983-2011 und 2011-2016 (5.2 und 5.3); ebenso zwischen allen applizierten verschiedenen prT einschließlich der verwendeten Gewebequellen im zeitlichen Verlauf 2001-2011 und 2011-2016 (5.3). Die Ergebnisse der Studien 1983-2011 sind dem metaanalytischen Teil der Studie von GEBUREK und STADLER (2011b) entnommen und überarbeitet.

Ein wichtiges Teilergebnis der Metaanalyse sind BehG, bei denen *GFR* empirisch ermittelt werden kann (5.4). Sie werden differenziert nach PrTB (5.4.2) und KtrB (5.4.3) untersucht mit dem Ziel poolbare inhaltlich homogene BehG zu finden. Dafür werden Vergleichskriterien (5.4.2.1) entwickelt. Mittels quantitativer Heterogenitätsanalyse wird der gepoolte Effektschätzer  $GFR_{\text{gepoolt}}$  (incl. Variabilität) ermittelt. Erscheint der Unterschied des interessierenden Effekts zwischen zwei BehG bedeutsam, so kann mittels Power-Analyse die Effektdifferenz  $d_{GFR}$  untersucht und per Signifikanznachweis statistisch gesichert werden, wenn die erforderliche Fallzahl gegeben ist. Das Ergebnis von Kapitel 5.4 wird in 5.4.5 zusammengefasst.

### 5.1 Behandlungsgruppen in den Studien 2011-2016

#### 5.1.1 Einflussfaktoren und interessierende Merkmale des Behandlungserfolgs

Im Diskurs mit der Forschung (2.2.2) können folgende maßgebliche Einflussfaktoren und Merkmale des Behandlungserfolgs eruiert und als Primärdaten definiert werden: Erkrankte Gewebestruktur, Lokalisation der behandelten Gliedmaße (vorne, hinten), Behandlungsform (KtrB /PrTB; appliziertes prT), Größe *n* der Studienpopulation, Anzahl *N* der behandelten Pferde einer extrahierten BehG, Nutzungsrichtung mit Angabe der Anzahl der Pferde, Erfolgsmerkmal EM, Erfolgsrate *E* (bzgl. EM), Rehabilitationsdauer, Prüfzeit zur Kontrolle des Erfolgsmerkmals auf ein Rezidiv, Rezidivrate *R*. Das häufigste Erfolgsmerkmal ist: 'Nach Rehabilitation kehren die Sportpferde wieder zur Nutzung auf dem ursprünglichen Leistungs-Niveau zurück'. Die Gewebestruktur wird nach 8 Lokalisationsmerkmalen unterschieden: (1) OBS, (2) TBS, (3) USB der TBS, (4) FT-Ursprung, (5) FT-Körper, (6) FT-Schenkel, (7) FT ohne Angabe der Lokalisation, (8) OBS-TBS-USB-FT gemischt. Aus den Studien werden die Primärdaten extrahiert und zu kleineren BehG als die Studienpopulation nach den Einflussfaktoren behandelte Gewebestruktur, Behandlungsform (PrTB, KtrB), appliziertes prT und Nutzungsart geordnet. Soweit möglich werden die quantitativen Anteilsraten *E* und *R* sowie die Effektgröße *GFR* (gemäß Modell) aus den Primärdaten einer Studie neu berechnet. Alle BehG werden mit ihren Primärdaten tabellarisch präsentiert: Tab 5.1a für das Sehnengewebe OBS, TBS, USB der TBS; Tab 5.1b für den FT-Ursprung, -Körper und -Schenkel. Aus Platzgründen wird nur der Erstautor notiert sowie das Erscheinungsjahr einer Studie. Bei der Gestaltung des Tabellendesigns lehnt sich die vorliegende Metaanalyse an die Studie von GEBUREK und STADLER (2011) an um den direkten Vergleich der Ergebnisse zwischen 2011-2016 und 1983-2011 zu erleichtern. Die Merkmale Rehabilitationsdauer und *GFR*, sowie in Tab 5.4 die Einflussfaktoren Alter der Pferde und Läsionsgrad kommen neu hinzu; außerdem werden die beiden Größen *n* bzw. *N* zwecks normierter Berechnungsweise von *GFR* unterschieden (5.5). Angemerkt sei: Die Ergebnisse der Studie Nr. 24 (Tab 5.1a und 5.1b) werden anhand der Dissertationsschrift LUTZ (2011) gemäß *GFR*-Modell für die einzelnen Gewebestrukturen neu berechnet. Die Ergebnisse der spanischen Studie ABELLANET (2009), die bei GEBUREK und STADLER (2011) nicht erscheint, werden in 2011-2016 gelistet.

## 5.2 Deskriptive Vergleiche zwischen den Studien im zeitlichen Verlauf 1983-2011 und 2011-2016

Die BehG werden statistisch deskriptiv bezüglich Anzahl, Behandlungsform, Größe und Variabilität der Patienten-Fallzahl unter verschiedenen Aspekten der quantitativen Verteilung verglichen und das Ergebnis präsentiert, wobei der zeitliche Verlauf 1983-2011 (GEBUREK und STADLER 2011) und 2011-2016 beachtet wird.

### 5.2.1 Häufigkeitsverteilung der BehG, Studien, Behandlungsformen (PrTB, KtrB) und Patienten in 1983-2011 und 2011-2016.

Tabelle 5.2.1 Häufigkeitsverteilung der BehG, Studien, Behandlungsformen (PrTB; KtrB) und Patienten in 1983-2011 und 2011-2016						
	1983-2001	2001-2011			2011-2016	
Behandlungsform	KtrB	KtrB	PrTB		KtrB	PrTB
Anzahl d. Studien	4	2	14	1*	6	24
Anzahl d. extrahierten BehG	6	3	21	1*	7	62
Anzahl d. behandelten Pferde	174	134	931	1000*	137	866
Summe d. behandelten Pferde	174	1065		1000*	1003	
* untypisch große BehG						

#### Studien-Anzahl, Patienten-Fallzahl und Behandlungsformen

Die erste große Studie, die Läsionen mit PrTB behandelt, stammt aus dem Jahr 2001 (HERTHEL 2001; Knochenmarkaspirat, n = 100). Vergleicht man die Zeiträume 2001-2011 und 2011-2016 (Tab 5.2.1), so reduziert sich die durchschnittliche Anzahl ca. 67 (1065 /16) behandelte Sportpferde pro Studie auf ca. 33 (1003 /30) um rund die Hälfte.

Auf den Zeitraum 1983-2001 entfallen 4 Studien (174 Pferde), die alle ausschließlich mit KtrB behandeln. Auf den Zeitraum 2001-2011 entfallen 17 Studien (ca. 2065 Pferde), 2 Studien zu KtrB (ca. 134 Pferde) und 14 Studien zu PrTB (ca. 931 Pferde). 1 Studie (der 17) zu PrTB mit ca. 1000 Pferde ist untypisch groß und differenziert das Lokalisationsmerkmal nicht „OBS, TBS, USB, FT gemischt“ (DAHLGREN 2009; Ad-M-SC); sie kann als Ausreißer betrachtet betrachtet werden.

Auf den Zeitraum 2011-2016 entfallen 30 Studien (ca. 1003 Pferde), 6 Studien (ca. 137 Pferde) zu KtrB und 24 Studien (ca. 866 Pferde) zu PrTB. 2 (der 30) Studien vergleichen beide Behandlungsformen (KtrB; PrTB).

#### BehG-Anzahl und Behandlungsformen

Da im Zeitraum 1983-2011 einige der 21 Studien mehr als eine Gewebestruktur untersuchen, entstehen bei der Extraktion 31 BehG (Tab 5.2.1) für insgesamt ca. 2239 (174 KtrB +2065 PrTB) behandelte Pferde. Grenzt man den Zeitraum auf 2001-2011 ein – da erst ab 2001 mit prT behandelt wird - so entfallen darauf 25 BehG aus 17 Studien mit ca. 2065 behandelten Pferden. Darunter ist wie erwähnt eine untypisch große BehG mit ca. 1000\* Pferden (Tab 5.2.1).

Da im Zeitraum 2011-2016 15 (der 30) Studien (Nr. 3, 4, 7, 8, 11, 12, 13, 14, 15, 16, 17, 21, 23, 24, 29) mehr als eine Gewebestruktur untersuchen oder in zwei BehG unterschiedliche prT oder konservative und regenerative Behandlungsform (KtrB; PrTB) vergleichen, entstehen 69 BehG (Tab 5.2.1) bezüglich insgesamt ca. 1003 behandelte Pferde.

Auf die Behandlungsform KtrB entfallen in 1983-2001 100 % (6 /6), in 1983-2011 29 % (9 /31) BehG, in 2001-2011 12 % (3 /25) BehG und in 2011-2016 10 % (7 /69) BehG. Auf die Behandlungsform PrTB entfal-

len entsprechend die komplementären Anteile: 1983- 2001 0 %, 1983-2011 71 %; 2001-2011 88 %, 2011-2016 90 % (Tab 5.2.1).

**Zwischenergebnis.** Der Anteil 10 % bzw. 12 % BehG zur Behandlungsform KtrB in 2001-2011 und 2011-2016 bleibt annähernd gleich niedrig. Für die Behandlungsform PrTB ist dagegen der entsprechend komplementäre Anteil 88 % bzw. 90 % gleichbleibend hoch.

Der niedrige Anteil der Behandlungsform KtrB steht in Kontrast zu der Erkenntnis, dass Behandlungserfolge, die mit KtrB erzielt werden, bei gleichen Bedingungen ein guter Vergleichsmaßstab für den Erfolg mit der Behandlungsform PrTB sein können. Das erklärt sich wahrscheinlich aus der seit 2001 wachsenden hohen Erwartungshaltung gegenüber prT.

### Patienten-Fallzahl und Behandlungsformen

Mit **KtrB** werden im Zeitraum 2011-2016 14 % (137 /1003) der Pferde behandelt. Im Zeitraum 2001-2011 sind dies 13 % (134 /1065). Wählt man den größeren Zeitraum 1983-2011 so wird ein sehr viel größerer Anteil 29 % (308 /1065) der Pferde mit KtrB behandelt. Denn in 1983-2001 kommen 4 Studien ausschließlich zu der Behandlungsform KtrB mit insgesamt 174 Pferden hinzu (308 = 134+174).

Auf die Behandlungsform **PrTB** entfallen entsprechend die komplementären Anteile: 1983-2011 71 %, 2001-2011 87 %, 2011-2016 86 % (Tab 5.2.1). - Wie erwähnt umfasst eine weitere untypisch große Studie 1000 der insgesamt 2065 Pferde in 2001-2011.

**Zwischenergebnis.** In 2011-2016 geht gegenüber 1983-2011 der Anteil der mit KtrB behandelten Pferde an der Gesamtzahl der behandelten Pferde von ca. 29 % auf ca. 14 % zurück, d. h. um rund die Hälfte. Der Rückgang zeigt sich bereits ab 2001-2011 (13 %) und fällt zeitlich mit dem Beginn der ersten großen Studie zur Behandlungsform PrTB in 2001 zusammen.

### 5.2.2 Variabilität der Patienten-Fallzahl in den BehG

	Anzahl BehG	Patienten-Fallzahl	Spannweite	Q <sub>1</sub>	Q <sub>3</sub>	Q <sub>3</sub> -Q <sub>1</sub>	Median	Mw	STD
1983-2011	30	1239	7-172	15	54	39	28	41	38
2011-2016	69	1003	1-85	5	20	15	9	15	14

BehG Behandlungsgruppe; Mw arithm. Mittelwert; Q<sub>i</sub> Quartil; STD Standardabweichung; Werte gerundet.

Im Zeitraum 1983-2011 streut die Fallzahl der 30 BehG mit der Spannweite 7-172 Pferde um den Median 28 Pferde. Die Quartile sind Q<sub>1</sub> = 15, Q<sub>3</sub> = 54 Pferde; Mw = 41; STD = 38. Der Inter-Quartils-Abstand beträgt IQR = Q<sub>3</sub>-Q<sub>1</sub> = 39 Pferde. Die 31ste BehG mit 1000 Pferden ist untypisch groß und bleibt hier außer Betracht. Im Zeitraum 2011-2016 streut die Fallzahl der 69 BehG mit der Spannweite 1–85 Pferde um den Median 9 Pferde. Die Quartile der Fallzahl betragen Q<sub>1</sub> = 5, Q<sub>3</sub> = 20 Pferde; Mw = 15; STD = 14. Der Inter-Quartils-Abstand beträgt IQR = 15 Pferde.

Der Vergleich der beiden Zeiträume ergibt: Die Inter-Quartils-Abstände IQR = Q<sub>3</sub> -Q<sub>1</sub> betragen 39 Pferde bzw. 15 Pferde und unterscheiden sich um rund das Zweieinhalbfache. Auf die beiden Mittelwerte, Mediane und Spannweiten trifft das Gleiche zu.

**Zwischenergebnis.** Die durchschnittliche Fallzahl 15 Pferde pro BehG der Studien in 2011-2016 ist gegenüber 41 Pferde der Studien in 1983-2011 weniger als halb so groß (Tab 5.2.2).



### 5.2.3 Verteilung der Patienten-Fallzahlklassen in den BehG

Tabelle 5.2.3 Verteilung der Patienten-Fallzahlklassen in den BehG							
1983-2011	30 BehG	Proz. Hfkt	0 %	26,7 %	26,7 %	33,3 %	13,3 %
		Absol. Hfkt.	0	8	8	10	4
Patientenfallzahl-Klasse			1-6	7-15	16-29	30-75	100-172
2011-2016	69 BehG	Absol. Hfkt.	29	18	12	10	0
		Proz. Hfkt.	42 %	26,1 %	17,4 %	14,5 %	0 %
Absol. Hfkt. = absolute Häufigkeit; BehG = Behandlungsgruppe; Proz. Hfkt. = prozentuale Häufigkeit							

In 1983-2011 verteilen sich die Patientenfallzahl-Klassen der 30 BehG wie folgt: 7-15 Pferde 26,7 % (8 /31) BehG; 16-29 Pferde 26,7 % (8 /31) BehG; 30-75 Pferde 33,3 % (10 /31) BehG; 100-172 Pferde 13,3 % (4 /31) BehG (zusätzlich 1 untypisch große BehG 1000 Pferde). In 2011-2016 verteilen sich die Patientenfallzahl-Klassen der 69 BehG wie folgt: 1-15 Pferde 68,1 % (47 /69) BehG; 16-29 Pferde 17,4 % (12 /69) BehG; 32-85 Pferde 14,5 % (10 /68) BehG (Tab 5.2.3).

**Zwischenergebnis.** Während 73,3 % der BehG aus den Studien in 1983-2011 die Fallzahl  $16 \leq n$  Pferde haben, hat ein fast ebenso großer Anteil, nämlich 68,1 % der BehG aus den Studien in 2011-2016  $n < 16$  Pferde. In 1983-2011 ist bei allen BehG die Fallzahl  $7 \leq n$  Pferde. Im Gegensatz dazu hat in 2011-2016 fast die Hälfte (42 %, 29/69) aller BehG die Fallzahl  $n < 7$  Pferde. Diese 42 % BehG umfassen nur 116 Pferde, d. h. 12 % (116 /1003) aller behandelten Pferde. Der abrupte Rückgang der Behandlungsgruppengröße ab 2011 hängt – wie sich (5.3.2) zeigen wird – wahrscheinlich mit der ab 2011 wachsenden Zahl neu entwickelter prT zusammen. Wodurch sich die Anzahl der BehG (30 vs. 69) rund verdoppelt, gleichzeitig aber die Behandlungsgruppengröße nahezu halbiert.

## 5.3 Vergleich der Gewebequellen und prT in den Studien 2001-2016

### 5.3.1 19 verschiedene prT

Im Folgenden werden in Tab 5.3.1 alle verschiedenen prT, die seit 2001 in den untersuchten Studien zur Behandlung von Erkrankungen des FT und Sehnenapparates bei Sportpferden appliziert werden, unter Angabe der Studien-Nummer (aus Tab 5.1a und 5.1b) präsentiert. Die Markierung mit einem Quadrat ■ bedeutet, dass die 'Gebrauchsfähigkeitsrate' *GFR* nicht angegeben werden kann. Teils weil die Erfolgsrate *E* oder die Rezidivrate *R* oder die Variablenwerte A, B, C zur Berechnung von *GFR* (5.5) nicht extrahierbar sind. Von den Studien in 2001-2011 (aus dem metaanalytischen Teil der Studie von GEBUREK und STADLER 2011b) werden nur solche mit ihrer laufenden Nummer aufgeführt, bei denen *GFR* ermittelt werden kann. Zusätzlich sind sie mit A gekennzeichnet um sie besser von Studien in 2011-2016 zu unterscheiden. Hierzu gehört u. a. die Studie 2A GODWIN et al. (2012), die unter ihrem offiziellen Erscheinungsjahr 2012 zitiert wird. Die übrigen Studien ohne *GFR* sind in ihrer Gesamtanzahl und mit Quadrat ■ notiert.

Vier Studien, Nr. 1, 2, 21, 28, fehlen hier in Tab 5.3.1, da sie ausschließlich BehG zur Behandlungsform KtrB beisteuern. Die beiden Studien Nr. 3, 4 dagegen, die in zwei BehG beide Behandlungsformen KtrB vs. PrTB vergleichen, werden in Tab 5.3.1 gelistet. - Aus Gründen der Übersichtlichkeit wird ein prT, das aus einer autologen bzw. allogenen Gewebequelle stammt, kurz mit „autologes prT“ bzw. „allogenes prT“ bezeichnet; Fettgewebe-MSC können auch mit Ad-MSC (adipose tissue derived MSC) und Knochenmark-MSC mit BM-MSC (bone marrow derived MSC) bezeichnet werden.

Tabelle 5.3.1 <b>19 verschiedene prT in Studien 2001-2011 und 2011-2016 mit Angabe der Studien-Nummer</b>					
Lf. Nr.		2001-2011	2011-2016		
	<b>PrT</b>	<b>autolog</b>			
1	Konditioniertes Plasma aus Knochenmark	1 Studie ■			
			<b>autolog</b>	<b>allogen</b>	<b>xenogen</b>
2	Knochenmark-MSC	3A, 2A, 4 Studien ■	11, 12 ■		
3	Fettgewebe-MSC	4A, 5A	25, 12 ■		
4	Knochenmarkaspirat	7A, 8A, 1 Studie ■	10 ■, 27 ■		
5	PRP (Filtration, Zentrifugation)	3 Studien ■	6, 7, 23, 24, 5■, 8■, 29■, 30■		
6	Knochenmark-MSC <u>und</u> PRP	1 Studie ■	13		
7	Azelluläre Matrix aus Urinblasengewebe vom Schwein				24
8	Konditioniertes Serum ACS		3		
9	Blut-MSC		9		
10	Fettgewebe-MSC <u>und</u> PRP		14		
11	Allogene Amniongewebe-MSC			11	
12	Konditioniertes Medium aus allogenen Amniongewebe-MSC			15, 16	
13	Allogene Knochenmark-MSC			30 ■	
14	Allogenes PRP			21 ■	
15	Allogene Nabelschnurblut-MSC			4 ■, 18 ■	
16	Sehnenzell-induzierte MSC aus allogenen Blut <u>und</u> allogenes PRP			26	
17	Sehnenzell-induzierte MSC aus allogenen Blut Tendo-Cell®			22 ■	
			<b>autolog und allogen</b>		
18	Allogene Fettgewebe-MSC <u>und</u> autologes PRP		17 ■		
19	Ovine Amnion-Epithel-Gewebe-MSC				19 ■, 20
■ = Angabe von <i>GFR</i> nicht möglich					

### 5.3.2 Verteilung der 19 verschiedenen prT differenziert nach autologer, allogener und xenogener Gewebequelle

Im zeitlichen Verlauf von 2001-2011 und 2011-2016 verteilen sich die 19 prT geordnet nach autologen, allogenen und xenogenen Gewebequellen wie folgt.

**Autologe Gewebequelle:** In 2001-2016 werden 9 verschiedene autologe prT appliziert. Nur 1 (von 9) prT, konditioniertes Plasma aus Knochenmark, wird ausschließlich im Zeitraum 2001-2011 eingesetzt (KÜNNKE 2008). Weitere 5 (von 9) prT werden sowohl im Zeitraum 2001-2011 als auch 2011-2016 eingesetzt: Knochenmark-MSC; Fettgewebe-MSC; Native Knochenmarkaspirate; Knochenmark-MSC kombiniert mit PRP. Die übrigen 3 (von 9) prT werden nur in 2011-2016 eingesetzt: Konditioniertes Serum; Blut-MSC; Fettgewebe-MSC kombiniert mit PRP.

**Allogene Gewebequelle:** Im Zeitraum 2011-2016 wird (erstmalig 2011) mit allogenen prT behandelt. Es sind insgesamt 7 verschiedene allogene prT: allogene Nabelschnurblut-MSC; allogene Amniongewebe-MSC; allogene Knochenmark-MSC; konditioniertes Medium aus allogenen Amniongewebe-MSC; allogenes PRP;

sehnenzell-induzierte MSC aus allogenen Blut kombiniert mit allogenen PRP; sehnenzell-induzierte MSC (Tendo-Cell®) aus allogenen Blut.

**Autologe und allogene Gewebequelle kombiniert:** Nur 1 prT (Nr. 17) kombiniert im Zeitraum 2011-2016 allogene Fettgewebe-MSC mit autologem PRP.

**Xenogene Gewebequelle:** In 2011-2016 sind es 2 xenogene prT: Nr. 19, Nr. 20 xenogene ovine Amnion-Epithel-Gewebe-Zellen und Nr. 24 porcine UBM, d. h. extrazelluläre Matrix aus Urinblasengewebe vom Schwein.

### 5.3.3 Zunehmende Diversifizierung der prT in den Studien im Verlauf 2001-2011 und 2011-2016

Im zeitlichen Verlauf 2001-2011 und 2011-2016 verteilen sich die Studien, die mit unterschiedlichen prT behandeln, wie folgt: Im Zeitraum 2001-2011 wird in insgesamt 16 Studien nur mit autologen prT behandeln. Es werden 6 verschiedene autologe prT (38 %; 6 /16) appliziert. Im Zeitraum 2011-2016 behandeln insgesamt 26 Studien mit prT. Es werden 19 verschiedene (73,1 %; 19 /26) prT eingesetzt, nämlich 9 autologe (47,4 %; 9 /19) prT, 7 allogene (36,8 % (7 /19) prT, 1 autolog-allogenes (5,3 %; 1 /19) Kombinations-prT und 2 xenogene (10,5 %; 2 /19) prT.

**Zwischenergebnis.** Im Zeitraum 2011-2016 steigt die Diversifizierungsrate der prT, d. h. die Zahl unterschiedlicher prT, die zur Behandlung von Erkrankungen eingesetzt werden, bezogen auf die Gesamtzahl der Studien, die mit prT behandeln, gegenüber 2001-2011 von 38 % (6/16) auf 73,1 % (19/26); das entspricht einer Zunahme um den Faktor ca. 1,9. Der in 2011-2016 verstärkte Trend neue prT zu untersuchen entspringt vor allem dem wachsenden Interesse an allogenen prT, die u. E. ab 2011 erstmals eingesetzt werden.

### 5.3.4 Verteilung der Studien differenziert nach autologen, allogenen, xenogenen prT

Da 5 Studien (Nr. 11, 12, 22, 24, 30) jeweils in zwei BehG zwei verschiedene prT vergleichen, übersteigt die Summe der Studienanteile 100 % (exakt 112 %). Nr. 11 und Nr. 30 vergleichen die Wirkung eines autologen mit einem allogenen prT. Nr. 24 vergleicht ein xenogenes mit einem autologen prT. Nr. 12 vergleicht die Wirkung von zwei autologen prT. Nr. 22 vergleicht die Wirkung von zwei allogenen prT. Außerdem vergleicht Nr. 3 die Wirkung eines autologen prT mit KtrB und Nr. 4 eines allogenen prT mit KtrB.

**Zeitraum 2001-2011** (Tab 5.3.1; 1. Spalte): 16 (der 21) Studien behandeln mit prT, und zwar alle nur mit autologen prT. 5 (der 21) Studien behandeln mit KtrB.

**Zeitraum 2011-2016** (Tab 5.3.1; 2. Spalte u. 3. Spalte): 26 (von 30) Studien behandeln mit prT; 4 (von 30) Studien mit KtrB. 65 % (17 /26) der Studien behandeln mit prT aus autologen Gewebequellen: Nr. 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 23, 24, 25, 27, 29, 30. 35 % (9 /26) behandeln mit prT aus allogenen Gewebequellen: Nr. 4, 11, 15, 16, 17, 18, 22, 26, 30. Nur 1 Studie 4 % (1 /26) verwendet ein Kombinations-prT aus einer allogenen und einer autologen Gewebequelle, Nr. 17. 12 % (3 /26) behandeln Erkrankungen mit prT aus xenogenen Gewebequellen: Nr. 19, Nr. 20 und Nr. 24.

**Zwischenergebnis.** Im Zeitraum 2001-2011 wird in den Studien nur prT aus autologen Gewebequellen verwendet (100 %; 16 /16). Im Zeitraum 2011-2016 kommen zu autologen erstmals allogene prT mit einem Studienanteil von 35 % (9 /26) hinzu (zuerst Nr. 30 in 2011); im gleichen Zeitraum geht der Studienanteil der autologen prT von 100 % (16 /16) auf 65 % (17 /26) zurück.

Bemerkenswert ist: 2013 erfolgt erstmals die Behandlung mit einem Substrat kombiniert aus einer allogenen und einer autologen Gewebequelle, Nr. 17, allogene Fettgewebe-MSC kombiniert mit autologem PRP (n = 19); diese Kombination bleibt ein Einzelfall. (Da auch *GFR* nicht ermittelt werden kann, ist das Ergebnis von

Nr. 17 nicht vergleichbar).

### 5.3.5 Verteilung der Studien differenziert nach Knochenmark, Fettgewebe, Amniongewebe und Blut als Gewebequelle für MSC sowie nach Blutprodukten und UBM

In 2001-2011 behandeln 16 Studien und in 2011-2016 26 Studien mit Knochenmark, Fettgewebe, Blut (incl. PRP) und Amniongewebe. 3 der 26 Studien (Nr. 11, 12, 24) vergleichen zwei verschiedene prT und weitere 4 der 26 Studien kombinieren PRP mit einer anderen Gewebequelle (Nr. 13, 14, 17, 26). Die Studienanteile verteilen sich im zeitlichen Verlauf wie folgt auf die Gewebequellen:

<b>Knochenmark</b>	Der Studienanteil für prT aus <b>autologem</b> Knochenmark fällt von 69 % (11 /16) in <u>2001-2011</u> auf 19 % (5 /26) in <u>2011-2016</u> . Darunter ist 1 Studie, die Knochenmark mit PRP kombiniert (Nr. 13, 2013). <b>Allogene</b> Knochenmark-MSC werden nur in <u>2011-2016</u> und nur in 1 Studie 4 % (1 /26; Nr. 30) eingesetzt.
<b>Fettgewebe</b>	Der Studienanteil für prT aus <b>autologem</b> Fettgewebe bleibt konstant 12,5 % (2 /16) in <u>2001-2011</u> und 12 % (3 /26) in <u>2011-2016</u> . Darunter ist 1 Studie, die Fettgewebe mit PRP kombiniert (Nr. 14, 2015). <b>Allogene</b> Fettgewebe-MSC werden nur in <u>2011-2016</u> und nur in einer Studie (Nr. 17) kombiniert mit autologem PRP 4 % (1 /26) eingesetzt.
<b>Amniongewebe</b>	Nur in <u>2011-2016</u> wird <b>allogenes</b> Amniongewebe als Gewebequelle für prT in 3 Studien 12 % (3 /26) und <b>xenogenes</b> (ovines) Amniongewebe in 2 Studien 8 % (2 /26) eingesetzt.
<b>Blut-MSC</b>	<u>2001-2011</u> wird weder <b>autologes</b> noch <b>allogenes Blut</b> als Gewebequelle für MSC eingesetzt. Im Zeitraum <u>2011-2016</u> verwendet nur 1 Studie (Nr. 9) <b>autologes</b> Blut als Gewebequelle für MSC 4 % (1 /26) und 4 Studien 15 % (4 /26) <b>allogenes</b> Blut: nämlich Nabelschnurblut-MSC (Nr. 4, Nr. 18); sehnenzell-induzierte Blut-MSC (Nr. 22, Tendo-Cell®); sehnenzell-induzierte Blut-MSC kombiniert mit allogenen PRP (Nr. 26).
<b>PRP</b>	In <u>2001-2011</u> erreicht <b>autologes</b> PRP den Studienanteil 19 % (3 /16; erstmals WASE LAU et al. 2008). In <u>2011-2016</u> ist <b>autologes</b> PRP mit 42 % (11 von 26) Studienanteil das am häufigsten applizierte prT. Neben 8 (von 26) Studien, die ausschließlich PRP allein applizieren, kombinieren 3 (von 26) Studien (Nr. 13, 14, 17) PRP mit MSC aus einer anderen Gewebequelle. Der Studienanteil für <b>allogenes</b> PRP beträgt 8 % (2 /26; erstmals BEERTS et al. 2013). Davon kombiniert 1 Studie (Nr. 26) PRP mit sehnenzell-induzierten allogenen Blut-MSC. Noch gibt es keine klinische Studie, die PRP mit MSC vergleicht (SCHNABEL 2016).
<b>ACS</b>	Autologes konditioniertes Serum wird nur in <u>2011-2016</u> und nur in 1 Studie appliziert 4 % (1 /26; Nr. 8); ebenso
<b>UBM</b>	extrazelluläre Matrix aus porcinem Urinblasengewebe 1 Studie 4 % (1 /26; Nr. 24).

**Zwischenergebnis.** PRP und Knochenmark-MSC repräsentieren den prozentual höchsten Studienanteil zur Behandlungsform PrTB in 2001-2016: Im zeitlichen Verlauf 2001-2011 und 2011-2016 steigt er für PRP von 19 % auf 42 %, für autologe Knochenmark-MSC geht er von 69 % auf 19 % zurück; für Ad-MSC bleibt er konstant (12 %; 12,5 %). Nur in 2011-2016 werden allogene Amniongewebe-MSC (12 %) und xenogene ovine Amniongewebe-MSC (8 %) eingesetzt. Zum Vergleich: Der Studienanteil zu KtrB fällt im zeitlichen Verlauf 2001-2011 und 2011-2016 von 29 % auf 20 %.

### 5.3.6 Absolute Anwendungshäufigkeit der insgesamt 9 Gewebequellen in Studien 2001-2016

Die prozentuale Verteilung der Studien nach verwendeter Gewebequelle für prT (5.3.5) wird durch die Verteilung der absoluten Anwendungshäufigkeit der 3 autologen, 4 allogenen sowie 2 xenogenen Gewebequellen in den Studien 2001-2016 (Tab 5.3.6) im tabellarischen Überblick wie folgt ergänzt:

Tabelle 5.3.6 Absolute Anwendungshäufigkeit der 3 autologen, 4 allogenen und 2 xenogenen Gewebequellen in Studien 2001-2016									
			Gewebequelle für MSC				Blutprodukte		UBM
			Knochenmark	Fett-Gewebe	Amnion-Gewebe	Blut	PRP	Konditioniertes Serum	–
Anzahl der Studien	2001-2011	autolog	11	3	–	–	3	–	–
	2011-2016	autolog	5	2	–	–	11	1	–
		allogen	1	1	3	4	2	–	–
		xenogen	–	–	2 x ovin	–	–	–	1 x porcin

### 5.4 Behandlungsgruppen mit Angabe der 'Gebrauchsfähigkeitsrate' *GFR*

Bei Extraktion der Primärdaten aus den 30 Studien in 2011-2016 und 21 Studien in 1983-2011 lässt sich nicht für alle BehG die empirische 'Gebrauchsfähigkeitsrate' *GFR* ermitteln. Da nur bei letzteren BehG die GF der Sportpferde verglichen werden kann, sind sie ein wichtiges Teilergebnis der vorliegenden Metaanalyse. In 1983-2001 behandelt noch keine große Studie mit prT, sondern nur mit KtrB; jedoch lässt sich für KtrB keine BehG mit *GFR* extrahieren.

Aus den bisherigen Untersuchungsergebnissen folgt: Aus den Studien in 2011-2016 lassen sich 69 BehG (Tab 5.1a und Tab 5.1b) und in 2001-2011 31 BehG, somit insgesamt 100 BehG extrahieren. Bei 45 (der 69) und 5 (der 31) BehG lässt sich *GFR* bestimmen; das entspricht 65 % (45 /69) bzw. 16 % (5 /31). Insgesamt ergeben sich somit in 2001-2016 50 BehG mit *GFR*. Bei den übrigen 24 (von 69) bzw. 26 (von 31) BehG ist entweder die Erfolgsrate *E* unterschiedlich definiert bzw. *E* oder *R* werden nicht berichtet. Teilweise lassen sich auch nicht alle Werte der drei Variablen A, B, C zur Berechnung von *GFR* (5.5) ermitteln.

Tab 5.4 präsentiert die Verteilung der 50 BehG mit *GFR* differenziert nach Behandlungsform (PrTB; KtrB) und Gewebestruktur. Zusätzlich wird (in entsprechend angeordneten Feldern) die zugehörige Patienten-Fallzahl angezeigt.

Tabelle 5.4 Verteilung der 50 BehG (2001-2016) mit Angabe der <i>GFR</i> differenziert nach Behandlungsform und Gewebestruktur: OBS, TBS, FT sowie Kombinationen									
Anzahl der extrahierbaren BehG		OBS	TBS	FT				OBS-FT gemischt	OBS-TBS-USB gemischt
				Ursprung	Körper	Schenkel	ohne Angabe der Lokalisation		
PrTB	2011-2016	11	4	6	6	8	6		
	2001-2011	1		0			1	1	2
KtrB	2011-2016	2	0	0	0	1			
	2001-2011	1		0					

Fortsetz. Tabelle 5.4 Anzahl der Patienten		OBS	TBS	FT				OBS-FT gemischt	OBS-TBS-USB gemischt
				Ursprung	Körper	Schenkel	ohne Angabe der Lokalisation		
<b>PrTB</b>	2011-2016	181	21	81	35	80	180		
	2001-2011	113	0	0	0	0	48		155
<b>KtrB</b>	2011-2016	15	0	0	0	58			
	2001-2011	46		0					

Die Untersuchung der OBS ist mit 15 (der 50) BehG am häufigsten belegt. Zu PrTB (46 BehG) lassen sich bei den Gewebestrukturen FT-Ursprung, FT-Körper sowie FT-Schenkel in 2011-2016 20 (von 46) BehG mit Angabe von *GFR* aus den Studien extrahieren, aber keine in 2001-2011. Auf Kombinationen der Gewebestruktur oder „ohne Angabe beim FT“ entfallen in 2011-2016 20 % (10 /46) BehG. Sie umfassen ca. 43 % (383 /894) der behandelten Patienten. Somit verbleiben für eine nach der Gewebestruktur differenzierte Untersuchung mit Angabe von *GFR* ca. 57 % (511 /894) aller mit prT behandelten Patienten.

Die 50 BehG mit Angabe von *GFR* behandeln mit 10 prT der insgesamt 19 (darunter 2 prT-Kombinationen) verschiedenen prT, die im Zeitraum 2001-2016 therapeutisch eingesetzt werden (Tab 5.3.1). Für die übrigen 9 (der 19) prT ergeben sich BehG nur ohne *GFR*. Die 2 prT-Kombinationen (unter den 19 prT) sind: Fettgewebe-MSC kombiniert mit PRP; Knochenmark-MSC kombiniert mit PRP.

#### 5.4.1 Arbeitsschritte der Analyse der 50 BehG mit Angabe der *GFR*

Für die 50 BehG werden folgende Merkmale notiert: Gewebestruktur, appliziertes prT, Nutzungsart, *GFR*, Alter der Pferde, Läsionsgrad, Vorder- bzw. Hintergliedmaße, Prüfzeit, Rehabilitationsdauer. Der Läsionsgrad wird nach RANTANEN et al. (2011) mit LG 1-6 klassifiziert. Bei wenigen Studien, die den Läsionsgrad davon abweichend klassifizieren, wird er anhand der verfügbaren Angaben vergleichbar neu eingeschätzt und mit LG\* 1-6 gekennzeichnet. PrT aus allogenen Gewebequellen werden als solche ausdrücklich bezeichnet. Dagegen werden prT aus autologen Gewebequellen zugunsten einer übersichtlichen Darstellung nicht extra gekennzeichnet. Die Abkürzungen aus den Haupttabellen eingangs, Tab 5.1a u. 5.1b, werden beibehalten.

##### (1) Arbeitsschritt: Heterogenitätsanalyse.

Bei den 50 BehG soll differenziert nach den beiden Behandlungsformen KtrB bzw. PrTB geprüft werden, inwieweit sie vergleichbar sind. Gemäß Überblick in Tab 5.4 ergeben sich für die **Behandlungsform PrTB** 46 (der 50) BehG. Sie lassen sich nach 2 Kategorien A und B ordnen. Kategorie A: Aus 22 (der 50) BehG können 11 Paare mit jeweils 2 BehG gebildet werden, so dass Gewebestruktur und prT paar-intern identisch sind. Die 11 Paare werden in Tab 5.4.2.2 präsentiert. Um sie auf Poolbarkeit untersuchen zu können, werden in 5.4.2.1 qualitative Vergleichskriterien definiert. Für 4 (der 11) Paare werden sie erfüllt und der resultierende Effektschätzer  $GFR_{\text{gepoolt}}$  mithilfe der quantitativen Heterogenitätsanalyse bestimmt. Kategorie B: Weitere 24 (der 50) BehG unterscheiden sich in Gewebestruktur und prT. Sie sind daher nicht poolbar und werden als '24 alleinstehende' BehG bezeichnet; ihre Merkmale werden in Tab 5.4.5.1 präsentiert. Gemäß Tab 5.4 verbleiben 4 (der 50) BehG für die **Behandlungsform KtrB**; sie werden in Tab 5.4.3 präsentiert. 3 (der 4) BehG zu KtrB können zu 1 BehG bei der OBS gepoolt werden.

##### (2) Arbeitsschritt: Power-Analyse

Nachdem die Heterogenitätsanalyse abgeschlossen ist, deutet die Effektdifferenz  $d_{GFR}$  zwischen 2 vergleichbaren BehG auf einen Unterschied des Behandlungserfolgs. Die Effektdifferenz  $d_{GFR}$  kann mittels Power-Analyse per Signifikanznachweis statistisch gesichert werden, wenn die erforderliche Fallzahl gegeben ist. Dazu wird vorausgesetzt, dass der paarweise Vergleich pV zwischen 2 BehG adäquat ist. Dies soll systema-

tisch anhand von 2 Fallkonstellationen geprüft werden.

Zwischen 2 BehG ist: (1) **die Gewebestruktur gleich, aber die Behandlungsform (PrTB; KtrB) verschieden**; oder (2) **die Gewebestruktur verschieden, aber die Behandlungsform gleich**. BehG, die sich weder (1) noch (2) zuordnen lassen, unterscheiden sich in Behandlungsform und Gewebestruktur; pV sind nicht sinnvoll. Außerdem scheiden BehG, die Gewebestrukturen kombinieren, grundsätzlich aus, da solche BehG nicht vergleichbar sind.

Theoretisch erhöht sich mit steigender Fallzahl N-A einer BehG die Chance für einen validen Schätzwert *GFR*. Für die Realisierung signifikanter Ergebnisse mittels Power-Analyse ist eine umso größere Fallzahl erforderlich, je kleiner die Effektdifferenz zwischen 2 BehG ist. Von den insgesamt 50 BehG mit *GFR* in 2001-2016 haben 13 (26 %; 13 /50) die Fallzahl  $30 \leq N-A$ .

#### 5.4.2 Heterogenitäts-Analyse zur Behandlungsform PrTB

##### 5.4.2.1 Kriterien für die Poolbarkeit von BehG

Die Daten zweier BehG erscheinen poolbar, wenn sie neben Gewebestruktur und prT in weiteren 4 maßgeblichen Einflussfaktoren – Nutzungsart, Läsionsgrad, Altersklasse und Prüfzeitdauer – identisch sind oder zumindest gut übereinstimmen. Für die erforderlichen Merkmale der Übereinstimmung werden qualitative Vergleichskriterien entwickelt. Unterschiede von erheblichem Ausmaß, bei denen das Poolen ausgeschlossen sein soll, werden soweit sachdienlich durch geeignete quantitative Parameterintervalle definiert.

1) **Gewebestruktur**. Gemäß *GFR*-Modell sind nur Daten von BehG derselben biologischen Gewebestruktur poolbar.

2) **Prüfzeit**. Das Poolen erscheint nicht sinnvoll, wenn sich die Prüfzeiten zweier BehG erheblich unterscheiden, da mit wachsender Prüfzeit die Wahrscheinlichkeit für ein Rezidiv unkontrolliert steigt. Unterschiede werden als erheblich eingestuft, wenn die längere Prüfzeit das 1,5-fache der kürzeren übersteigt; d. h. 50 % größer ist als die kürzere (bspw. 12 M vs. 18 M oder mehr). Häufig begegnet in den Studien 12 vs. 24 M, was nicht poolbar ist.

3) **Läsionsgrad**. Der Läsionsgrad ist gemäß der Einteilung von RANTANEN et al. (2011) auf der Grundlage von ES, FAS und CSA semiquantitativ und quantitativ messbar (Tab 5.6.3). Höher gradige Verletzungen bei Sehngewebe haben eine schlechtere Heilungsprognose (MARR et al. 1993; GEBUREK et al. 2016). Werden Erkrankungen unterschiedlicher Läsionsgrade mit demselben prT behandelt, führt der biologische Heilungsprozess daher wahrscheinlich zu qualitativ unterschiedlichen Ergebnissen. Das resultierende Ersatzgewebe neigt unter gleicher mechanischer Belastung bei einer schweren Läsion eher zu einem Rezidiv als bei einer leichten. Erhöht sich die Rezidivrate *R*, verringert sich *GFR*. Erhebliche Unterschiede werden plausiblerweise durch drei per US-Diagnostik gut unterscheidbare Abstufungen quantifiziert: (1) minimale Läsionen LG 1-3; (2) leichte bis mittelschwere Läsionen LG 4-5; (3) schwere Läsion LG 6.

4) **Altersklasse**. Während der Rehabilitationsphase werden die Sportpferde individuell durch ein kontrolliertes Bewegungsprogramm, das an die Nutzungsart, das Alter, den Läsionsgrad sowie die sportliche Kondition angepasst ist, wieder an das ursprüngliche Leistungsniveau herangeführt. Die Patienten sollen das Maß an Kraft, Ausdauer, Flexibilität, Schnelligkeit und Kondition möglichst wieder erreichen wie vor Behandlung. Das erworbene Maß hängt einerseits mit angeborenen bzw. genetisch festgelegten Eigenschaften zusammen, andererseits mit der motorischen Entwicklung und der Fähigkeit der Bewegungskoordination durch Training. Beispielsweise beginnt bei Cutting Horses das Training bereits im Alter von 18-24 Monaten (TIPTON et al. 2013). Bei früh trainierten Pferden reduziert sich mittels hypertrophischer Antwort die Inzidenz von Sehnen- und Bänderverletzungen (KASASHIMA et al. 2002). Biologische Gewebe und konditionelle Fähigkeiten un-

terliegen durch Alterung einem natürlichen Abbauprozess. Wie erwähnt (2.2.2) deuten klinisch diagnostische, histomorphologische und zellbasierte experimentelle Untersuchungsergebnisse darauf, dass alterndes Bindegewebe im Vergleich zu jungem weniger reißfest wird, nach Insult langsamer heilt und ein schlechteres Ersatzgewebe bildet. Dadurch steigt natürlicherweise die Wahrscheinlichkeit für ein Rezidiv. Daher liegt es nahe, die Untersuchung der Gebrauchsfähigkeit von Sportpferden gemäß *GFR*-Modell nach Altersklassen zu differenzieren um Störfaktoren auf die quantitative *GFR* zu vermeiden. Erhebliche Unterschiede lassen sich plausiblerweise durch die Abstufung zwischen der Altersklasse bis ca. 16 Jahre und höheren Alters quantifizieren. Zwar kann feiner unterschieden werden, nämlich bis ca. 8 Jahre, ca. 8-16 Jahre und älter, was jedoch beim Poolen der Daten hier entfallen muss, da die überwiegend kleinen BehG sich diesbezüglich heterogen zusammensetzen und noch kleiner würden. Was es nicht erlaubt präzise zu klären, wie groß der Einfluss der Altersklasse bis 8 Jahre auf *GFR* ist.

5) **Nutzungsart.** Der Bewegungsablauf während spezifischer Trainingsübungen oder Wettkampfanforderungen belastet bei Sportpferden bestimmte Gewebeabschnitte der Sehnen und Bänder biomechanisch auf charakteristische Weise (2.2.2). Daher haben wahrscheinlich unterschiedliche Nutzungsarten eine unterschiedliche Prävalenz für typische Verletzungen überdurchschnittlich belasteter Gewebestrukturen (SOUSA et al. 2017). Beispielsweise beobachten PLEVIN und MCLELLAN (2014) bei Thoroughbred Jährlingen ( $n = 85$ ), dass Verletzungen der FT-Schenkel überwiegend (62,4 %; 53 /85) an der linken Vordergliedmaße auftreten. Dies kann mit ihrem Training gegen den Uhrzeigersinn zusammenhängen oder mit den engen Wendungen, die auf nordamerikanischen Rennplätzen für die Pferde üblich sind. Da nach einer Verletzung das Ersatzgewebe nicht die ursprüngliche physiologische Qualität zurück erlangt (reparatio, keine restitutio ad integrum), ist die Wahrscheinlichkeit für ein Rezidiv natürlicherweise erhöht, wenn die sportlichen Leistungsanforderungen gleich bleiben. Betrachtet man daher ein und dieselbe Gewebestruktur bei unterschiedlichen Nutzungsarten, so beobachtet man mit DYSON (2004) bei z. B. an Tendonitis der OBS erkrankten Sportpferden in der Prüfzeit 24 M die Tendenz zunehmender Rezidivraten: Springen (23 %; 3/13), Endurance (25 %; 2/8), Dressur (42,8 %; 3/7), Eventing (43 %; 34/79), Hindernis-Rennpferde (National Hunt; 56 %; 14/25), Flachrennpferde (66,7 %; 2/3). Allerdings kann die Studie keinen signifikanten Unterschied zwischen den Rezidivraten nachweisen ( $p = 0,35$ ). Wahrscheinlich weil die Fallzahl 3, 7, 8, 13, 25, 79 (nur 25 und 79 sind größer) überwiegend klein ist. Jedoch rechtfertigt die speziell bei der OBS beobachtete Tendenz folgende allgemeine Annahme: Effekte auf *GFR*, die mit der Belastung ein und derselben Gewebestruktur bei verschiedenen Nutzungsarten mittels Rezidivrate zusammenhängen, sollten sich zwischen zwei BehG gleich stark quantitativ auswirken, wenn die Nutzungsrichtung zwischen ihnen homogen verteilt ist. Die Gefahr, dass eine heterogene Verteilung der Nutzungsart zwischen zwei BehG unkontrolliert Störfaktoren auf die quantitative Effektgröße *GFR* erzeugt, erscheint beim Poolen um so größer je kleiner die BehG sind. Als quantitatives Kriterium, bei dem das Poolen (eine genügend große Fallzahl in beiden BehG vorausgesetzt) nicht sinnvoll erscheint, lässt sich somit festhalten: Ein erheblicher Unterschied der Zusammensetzung der Nutzungsart zwischen zwei BehG besteht, wenn sich die Verteilung der Häufigkeit der sportlichen Disziplin zwischen ihnen signifikant unterscheidet.

6) **Störfaktoren** haben vielseitige Ursachen. Etwa indem die Sportpferde infolge einer Fehldiagnose nach der Behandlung zu früh aus der Rehabilitationsphase in die Prüfzeit entlassen werden oder in der Prüfzeit die sportliche Leistung zu rasch forciert wird (z. B. Trainerehrgeiz). Auch kann das Interieur bzw. der Charakter einzelner Patienten – etwa der ausgeprägte Bewegungsdrang – die Einhaltung des kontrollierten Bewegungsprogramms gefährden. Solche Fehlerquellen sollten entdeckt und im Zweifelsfall die Patienten rechnerisch eliminiert, d. h. gemäß *GFR*-Modell zur Fallgruppe A (5.5.3.1-2) gezählt werden um einen Bias zu vermeiden.



### 5.4.2.2 Heterogenitätsanalyse der 11 Behandlungsgruppen-Paare zu PrTB

Aus 22 (der 50) BehG können 11 BehG-Paare gebildet werden, so dass Gewebestruktur und appliziertes prT paar-intern übereinstimmen; die 11 Paare werden in Tab 5.4.2.2 präsentiert. Die Anzahl der Paare verteilt sich wie folgt auf die Gewebestruktur: OBS 3 Paare, TBS 0 Paare, FT-Ursprung 1 Paar, FT-Körper 1 Paar, FT-Schenkel 3 Paare, FT ohne Angabe der Lokalisation 2 Paare, OBS-FT gemischt 1 Paar.

Verteilung der Fallzahlklassen: 6 (der 22) BehG haben die Fallzahl  $30 \leq N-A$ ; dies sind 6 BehG (der wie erwähnt 13 unter den 50) BehG mit  $30 \leq N-A$ ; ihre Nummern sind in Tab 5.4.2.2 grafisch hervorgehoben. Bei weiteren 9 (der 22) BehG ist  $N-A \leq 10$ ; bei 1 (der 22) BehG ist  $N-A = 13$ ; bei 6 (der 22) BehG ist  $19 \leq N-A \leq 24$ .

Tabelle 5.4.2.2 11 Behandlungsgruppen-Paare der Behandlungsform PrTB

Legende: [M] Monate, [a] Jahre, P Prüfzeit; LG Läsionsgrad, v vorne, h hinten, ~ gleiche Angabe.

Studien-Nr /prT /Nutzung /Fallzahl	GFR	95 %-CI	Reha [M]	Prüfz. [M]	Alter [a]	LG	v /h	Studie
<b>OBS</b>								
<b>1P Knochenmark-MSK</b>								
	<b>GFR<sub>gepoolt</sub> = 68,8 % [0,61-0,76]</b>							
<b>2A</b> FL 8 HR 105	80 /113	70,8 % [0,61-0,79]	k.A.	24 < P	Mw 10	4-6 *	k. A.	GODWIN et al. 2012
<b>Nr. 11</b> D8 S9 FL7 E5 Trab3	20 /32	62,5 % [0,44-0,79]	4-12	24	2-12	5-6	v32	LANGE-C. et al. 2013a
(Nr. 11*)	25 /32	78 % [0,60-0,91]	4-12	24	2-12	5-6	v32	LANGE-C. et al. 2013a )
<b>2P PRP GFR<sub>gepoolt</sub> = 60,4 [0,47-0,73]</b>								
<b>Nr. 7</b> Trab35 S7 R2	27 /44	61,4 % [0,46-0,76]	3,5-5,5	12-42	k.A.	3-6	k.A.	ABELLANET et al. 2009
Nr. 6 D5 R4 E1	4 /7	57,1 % [0,18-0,90]	6	24	4-20 Mw12	4-5*	v10	GEBUREK et al. 2016
<b>3P Konditioniertes Medium aus allogenen Amniongewebe-MSK</b>								
Das Poolen der Daten ist nicht sinnvoll, da sich bei gleich großer kleiner Fallzahl die Prüfzeit erheblich unterscheidet.								
Nr. 15 D2 S3	4 /5	80 % [0,28-0,99]	4-15	24	5-12	4-5	v4 h1	LANGE-C. et al. 2013b
Nr. 16 D2 S3	4 /5	80 % [0,28-0,99]	6-10	12	5-12	4-5	v5	TASSAN et al. 2014
<b>FT Ursprung</b>								
<b>4P PRP</b> Das Poolen der Daten erscheint nicht sinnvoll, da Prüfzeit und Läsionsgrad erheblich unterschiedlich sind.								
Nr. 23 Spo 20	16 /20	80 % [0,56-0,94]	3-6	6-24	3-14 Med 5	4-5*	v20	ROMAGNOLI et al. 2015
Nr. 24 D4 S5 Trab1 V4	7 /12	58,3 % [0,30-0,86]	2-12	17 < P	3-24 Mw 10,5	3-4*	k. A.	LUTZ 2011
<b>FT Körper</b>								
<b>5P Konditioniertes Medium aus allogenen Amniongewebe-MSK</b>								
Das Poolen der Daten ist nicht sinnvoll, da sich bei gleich großer kleiner Fallzahl die Prüfzeit erheblich unterscheidet.								
Nr. 15 End2	2 /2	100 % [0,16-1,0]	4-15	24	6	4	v2	LANGE-C. et al. 2013b
Nr. 16 E2	2 /2	100 % [0,16-1,0]	6-10	12	6; 7	4	v2	TASSAN et al. 2014
<b>FT Schenkel</b>								
<b>6P Konditioniertes Medium aus allogenen Amniongewebe-MSK</b>								
Das Poolen der Daten ist nicht sinnvoll, da bei etwa gleichgroßer Fallzahl die Prüfzeit erheblich unterschiedlich ist.								
Nr. 16 D2 E1 V1 S2	4 /5	80 % [0,28-0,99]	4-15	12	6-15	4-6	v3 h3	TASSAN et al. 2014
Nr. 15 D2 End1 Vo1 S1	3 /4	75 % [0,33-1,0]	6-10	24	6-15	4-6	v3 h2	LANGE-C. et al. 2013b

Tabelle 5.4.2.2 Fortsetzung

Studien-Nr /prT /Nutzung /Fallzahl	<i>GFR</i>	95 %-CI	Reha [M]	Prüfz. [M]	Alter [a]	LG	v /h	Studie
<b>7P PRP</b>	<b><i>GFR</i><sub>gepoolt</sub> = 63,7 % [0,48-0,78]</b>							
Nr. 24 D6 S5 Trab5 V6 14 /22	64 %	[0,41-0,83]	2-12	17 <P	3-24	3-4*	medial	k. A. LUTZ 2011
Nr. 24 D11 S6 Trab1 V2 13 /20	65 %	[0,41-0,85]	2-12	~	~	~	lateral	k. A. LUTZ 2011
<b>8P Xen. UBM</b>	<b><i>GFR</i><sub>gepoolt</sub> = 75,9 % [0,49-0,95]</b>							
Nr. 24 D1 Trab2 V3 S1 6 /7	86 %	[0,42-1,0]	2-12	17 <P	3-24	3-4*	medial	k. A. LUTZ 2011
Nr. 24 D3 2 /3	67 %	[0,09-0,99]	2-12	~	~	~	lateral	k. A. LUTZ 2011

### Kombinationen verschiedener Gewebestrukturen

#### FT ohne Angabe der Lokalisation

<b>9P PRP</b>	Das Poolen der Daten ist nicht sinnvoll, da die Gewebestruktur des FT nicht differenziert wird. Angaben zu Läsionsgrad und Alter fehlen. Die Verteilung der Häufigkeit der 4 Nutzungsrichtungen D, Trab, S, V ist signifikant unterschiedlich; $\chi^2 = 9,520$ ; $df = 3$ ; $p = 0,023$ .							
Nr. 7 D2 Trab9 S5 V3 17 /19	89 %	[0,67-0,99]	3,5-5,5	12-42	k.A.	3-6	k.A.	ABELLANET et al. 2009
Nr. 24 D28 Trab9 S19 V15 45 /70	64 %	[0,52-0,75]	2-12	17 <P	3-24	3-4*	k.A.	LUTZ 2011
<b>10P Knochenmarkspirat</b>	Das Poolen der Daten ist nicht sinnvoll, da die Gewebestruktur des FT nicht differenziert wird. Die Prüfzeit ist erheblich unterschiedlich. Angaben zu Läsionsgrad und Alter fehlen.							
7A D S 24 20 /24	83 %	[0,63-0,95]	k.A.	6-18	3-15	4-5*	v17 h7	ROSENBR. et.al. 2004b
8A V 24 17 /24	71 %	[0,49-0,87]	k.A.	6	k.A.	k.A.	k.A.	ROSENBR. et.al. 2004a

#### OBS-FT gemischt

<b>11P Fettgewebe-MSC</b>	Das Poolen der Daten ist nicht sinnvoll, da die Gewebestrukturen OBS und FT nicht unterschieden werden. Die Prüfzeit ist erheblich unterschiedlich. Angaben zu Läsionsgrad und Alter fehlen.							
5A Trab 53 27 /53	51 %	[0,37-0,65]	k.A.	24	k.A.	k.A.	k.A.	LEPPÄNEN et al. 2009b
4A Trab 31 23 /31	74 %	[0,55-0,88]	k.A.	12	k.A.	k.A.	k.A.	LEPPÄNEN et al. 2009a

#### 5.4.2.2.1 Bei 4 der 11 Paare sind beide BehG paar-intern poolbar

**Überblick:** Bei 4 (der 11) Paaren, 1P, 2P, 7P, 8P, erscheint das Poolen der beiden BehG adäquat. Da bei 1P, 2P und 7P durch Poolen eine vergleichsweise große Fallzahl ( $N-A = 145$ ; 51; 42) entsteht, bieten sie potenziell aussagekräftige Ergebnisse für *GFR*. Bei 1P stimmt die Prüfzeit 24 M mit dem *GFR*-Modell überein. Unter den 11 Paaren sind 6 BehG mit der Fallzahl  $30 \leq N-A$ ; 1P und 2P schließen 3 (dieser 6) BehG ein; diese 3 BehG sind aus Studie 2A, Nr. 7, Nr. 11. Die 4 Paare behandeln mit 3 verschiedenen prT, nämlich: PRP (2P, 7P), Knochenmark-MSC (1P) und xenogene porcine UBM (8P).

#### Paar 1P OBS, Knochenmark-MSC

Das Poolen der Daten erscheint adäquat. Läsionsgrad bzw. Altersklasse unterscheiden sich leicht. *GFR*<sub>gepoolt</sub> = 68,8 % ist durch die Nutzungsart Hindernis-Rennen (72 %; 105 /145) geprägt. Die Prüfzeit ist wie im *GFR*-Modell 24 M. Paar 1P erreicht von allen 11 Paaren die größte gepoolte Fallzahl  $N-A_{\text{gepoolt}} = 145$  Pferde.

**Begründung:** Die erste BehG (2A; GODWIN et al. 2012) hat *GFR* = 70,8 %. Es ist  $N = N-A = 113$  Sportpferde ( $A = 0$ ); 105 HR sowie 8 FL; sie gehören zur mittleren Altersklasse ( $Mw = 10$  a). Die Verletzungen entstehen durch Überbeanspruchung und sind leichten bis schweren Grades LG 4-6\*. Ein gewisses Biasrisiko besteht laut Studie 2A insofern, als die Einhaltung des Bewegungsprogramms nicht überprüft wird. Die zweite BehG Nr. 11 (LANGE-CONSIGLIO et al. 2013a) hat  $N = N-A = 32$  ( $A = 0$ ) Sportpferde unterschiedlicher Nutzungsart (D8, S9, FL 7, E5, Trab3). Sie umfasst etwas jüngere Tiere (2-12 a). Die Verletzungen

sind mittleren bis schweren Grades, 18-mal LG 5, 14-mal LG 6, alle an den Vordergliedmaßen lokalisiert. Die Verteilung der Häufigkeit der Nutzungsrichtung zwischen beiden BehG ist signifikant unterschiedlich ( $\chi^2$ ;  $p < 0,0001$ ). Letzteres spricht jedoch nicht zwingend gegen die Poolbarkeit, da der Einfluss der Rennpferde insgesamt (128 der 145) weit überwiegt. Die größere zweite BehG (113) besteht nur aus Rennpferden und 15 (der 32) Pferde der kleineren BehG sind ebenfalls Rennpferde. Die Autoren von Studie Nr. 11 diskutieren das potenzielle Risiko für Detection Bias. Einerseits da in der Studienpopulation athletische Disziplin (D8, S9, FL 7, E5, Trab3) und Läsionsgrad (LG 5-6) erheblich variieren; andererseits da (1) für den optimalen Zeitpunkt der Applikation (nach akuter Inflammation vor Narbenbildung) und (2) für die optimale Anzahl der MSC geeignete Standards noch fehlen. In der Studien reicht die Spannweite ca. von  $1 \times 10^6$  bis  $50 \times 10^6$  Zellen pro ml; VERONESI et al. (2016) berichten sogar die Spannweite  $4,5 \times 10^4 - 3,8 \times 10^8$  Zellen pro ml in Studien 2004-2014.

**Besonderheit:** Bei der zweiten BehG können nicht alle erforderlichen Daten für die Berechnung von *GFR* aus der Studie extrahiert werden. *GFR* lässt sich aber quantitativ abschätzen. Es ist entweder  $GFR_1^* = 62,5\%$  oder  $GFR_2^* = 78\%$ . Sachlich erscheint es nicht gerechtfertigt beim Poolen der Daten vom  $Mw = 70,5\%$  auszugehen. Vielmehr müssen beide Anteilswerte *GFR* getrennt betrachtet werden. Für den Behandlungserfolg mit BM-MSK bei Läsionen der OBS beträgt somit der resultierende Effektschätzer entweder  $GFR_{I^*_{gepoolt}} = 68,8\%$  oder  $GFR_{2^*_{gepoolt}} = 72\%$ . Das „schlechtere“ der beiden Ergebnisse wird zugelassen.

**Heterogenitätsanalyse:** Die Heterogenitätsanalyse führt zu dem Ergebnis:  $I^2 = 0\%$  bzgl. RE und  $Q = 0,821$  mit  $p = 0,3625$  bzgl. FE. Für  $I^2$  ist das 95 %-CI =  $[0\% - 0\%]$ .  $I^2$  und  $p$  deuten auf 'nicht heterogen' (4.2.3.5). Das Poolen gemäß FE erscheint rechnerisch adäquat. Die resultierenden Effektschätzer bezgl. RE und FE sind identisch  $GFR_{gepoolt} = 68,76\%$ . Das 95 %-CI für  $GFR_{gepoolt}$  bzgl. RE ist  $[61,05\% - 75,97\%]$  bzgl. FE  $[60,60\% - 76,14\%]$ ; beide Intervalle unterscheiden sich praktisch nicht.

## Paar 2P OBS, PRP

Das Poolen erscheint nur eingeschränkt sinnvoll.  $GFR_{gepoolt} = 60,4\%$  ist durch die Nutzungsart Traber (ca. 69 %; 35 /51) geprägt. Es ist  $N - A_{gepoolt} = 51$  Pferde ( $A_{gepoolt} = 3$ ). In der ersten BehG ist für die Untergruppe der Traber  $GFR_{Traber} = 54,3\%$  (19 /35)  $[37,8 - 70,8\%]$ . Bezüglich Läsionsgrad und Prüfzeit bestehen deutliche Unterschiede zwischen beiden BehG.

**Begründung:** Die erste BehG (Nr. 7; ABELLANET et al. 2009) hat  $GFR = 61,4\%$ ; es ist  $N = N - A = 44$  ( $A = 0$ ; Trab35, S7, R2). Eine Altersangabe der Pferde fehlt. PRP wird meistens zweimal (im Abstand von ca. 15 Tagen) appliziert, vereinzelt öfter. Die Pferde haben akute oder chronische Verletzungen; wobei die Studie keinen signifikant unterschiedlichen Effekt auf den Behandlungserfolg feststellt. Der Läsionsgrad LG 3-6 streut von minimal bis schwer. Die Prüfzeit 12-42 M variiert erheblich. Studien-interne Biasrisiken werden von den Autoren nicht diskutiert.

In der zweiten BehG (GEBUREK et al. 2016, RCT) ist  $GFR = 57,1\%$ . Es sind  $N - A = 7$  Nicht-Rennpferde ( $N = 10$ ;  $A = 3$  Pferde; D5, R4, E1) der mittleren Altersklasse (4-20 a;  $Mw = 12$  a). Die Prüfzeit ist 24 M. Die Pferde haben Verletzungen leichten bis mittleren Grades LG 4-5\*. Bei Behandlung sind 7 (der 10) Verletzungen 1-3, die übrigen 3 sind 4, 7, 8 Wochen alt. 5 (der 10) Verletzungen der OBS sind marginal bzw. peripher; dies erscheint für die Nutzungsart Dressur typisch und kommt wahrscheinlich von der exzessiv seitlichen Belastung oder 'self-kicking'. 3 (der 10) Verletzungen erscheinen diffus, 2 (der 10) sind 'core lesions'. Laut Studie besteht ein Biasrisiko insoweit, da die diagnostische Aussagekraft der angewandten B-Mode US-Technik bei heterogenem Läsionstyp begrenzt ist. Die sensitivere UTC Technik wird im erforderlichen Umfang nicht eingesetzt. Ein Performance Bias-Risiko besteht nach Meinung der Autoren auch darin, dass das biologisch optimale Verhältnis zwischen Thrombozyten und Leukozyten zur Behandlung von Tendinopathien mit PRP noch unbekannt ist. Es variiert wahrscheinlich mit dem Heilungsstadium der Sehnenverletzung und beeinflusst vermutlich das klinisch therapeutische Ergebnis. Die fehlenden Standards erschweren die

Vergleichbarkeit klinischer Studien, da das Verhältnis bei verschiedenen Herstellungsverfahren für PRP (Filter; Zentrifugation etc.) stark variiert.

Die Verteilung der Häufigkeit der Nutzungsrichtung (Trab35, S7, R2 vs. D5, R4, E1) unterscheidet sich zwischen beiden BehG signifikant ( $\chi^2$ ;  $p < 0,0001$ ). Was aber bei der relativ kleinen Fallzahl ( $N-A = 7$ ), die beim Poolen in die sehr viel größere ( $N-A = 44$ ) erste BehG einfließt, auch deshalb kaum ins Gewicht fallen sollte, da diese 7 Pferde alle Nicht-Rennpferde (D, R incl. Freizeitpferde, E) sind. So dass der Einfluss der größeren ersten BehG mit der breiteren Streuung sowohl bei Läsionsgrad LG 3-6 als auch Prüfzeit 12-42 *M* dominiert.

**Heterogenitätsanalyse:** Die Untersuchung der beiden *GFR* auf Heterogenität ergibt bzgl. RE  $I^2 = 0 \%$  [0-0 %], bzgl. FE  $Q = 0,06498$  mit  $p = 0,7988$ . Beide Werte deuten auf 'nicht heterogen' (4.2.3.5). Somit erscheint das Poolen gemäß FE rechnerisch adäquat. Die resultierenden Effektschätzer bezgl. RE und FE sind identisch  $GFR_{\text{gepoolt}} = 60,39 \%$ . Das 95 %-CI für  $GFR_{\text{gepoolt}}$  bzgl. RE ist [47,01 %-73,03 %] bzgl. FE [46,02 %-73,56 %]. Beide Intervalle unterscheiden sich praktisch nicht.

#### **Paar 7P FT-Schenkel, PRP**

Das Poolen der Daten erscheint adäquat. Es ist  $GFR_{\text{gepoolt}} = 63,7 \%$ ; ( $N-A$ )<sub>gepoolt</sub> = 42 Pferde. Bezüglich Läsionsgrad und Prüfzeit bestehen keine, bezüglich der Nutzungsart minimale Unterschiede. Jedoch wird die Altersklasse ab 16 Jahre nicht gesondert untersucht.

**Begründung:** Beide BehG gehören zur selben Studie (LUTZ 2011) und unterscheiden sich nur bezüglich der Lokalisation der Verletzung am medialen bzw. lateralen FT-Schenkel. Dressurpferde 40 % (17 /42) überwiegen; es folgen Springpferde 26 % (11 /42), Traber 14 % (6 /42) und 20 % sonstige sportliche Nutzungsrichtungen (auch Freizeitpferde). Die Verletzungen LG 3-4 sind minimalen bis leichten Grades. Zwei Drittel der Patienten sind akut erkrankt. Die Prüfzeit beträgt mindestens 17 *M*. Alle Altersklassen (3-24 *a*;  $Mw = 10$  *a*) sind vertreten. Die Verteilung der Häufigkeit der Nutzungsrichtung unterscheidet sich zwischen beiden Gruppen (D6, S5, Trab5, V6 vs. D11, S6, Trab1, V2) nicht signifikant ( $\chi^2$ ;  $p = 0,105$ ). Biasrisiken werden vom Autor nicht diskutiert.

**Heterogenitätsanalyse:** Die Untersuchung der beiden *GFR* auf Heterogenität ergibt:  $I^2 = 0 \%$  [0-0 %] bzw.  $Q = 0,007377$  mit  $p = 0,9316$ . Beide deuten auf 'nicht heterogen' (4.2.3.5). Das Poolen gemäß FE erscheint somit rechnerisch adäquat. Die resultierenden Effektschätzer bezgl. RE und FE sind identisch  $GFR_{\text{gepoolt}} = 63,653 \%$ . Das 95%-CI für  $GFR_{\text{gepoolt}}$  bzgl. RE ist [47,05 %-77,07 %] bzgl. FE [47,79 %-77,61 %]. Beide Intervalle unterscheiden sich praktisch nicht.

#### **Paar 8P FT Schenkel, Xenogene porcine UBM**

Das Poolen der Daten ist möglich; es ist  $GFR_{\text{gepoolt}} = 75,9 \%$ .

**Begründung:** Die Fallzahl  $N-A_{\text{gepoolt}} = 10$  (= 7+3) Sportpferde ist sehr klein. Alle Altersklassen (3-24 *a*) sind vertreten. Die Altersklasse der Patienten ab 16 Jahre wird allerdings nicht gesondert untersucht. Die Läsionen sind minimalen bis leichten Grades LG 3-4\*. Die Häufigkeit der Nutzungsrichtung (D1, Trab2, S1, V3 vs. D3) verteilt sich signifikant unterschiedlich ( $\chi^2$ ;  $p < 0,0001$ ). Wie bei Paar 7P gehören beide BehG zu derselben Studie (LUTZ 2011) und unterscheiden sich nur bezüglich der Lokalisation der Verletzungen am medialen bzw. lateralen FT-Schenkel.

**Heterogenitätsanalyse.** Die Untersuchung von  $GFR_1 = 85,7 \%$  [0,42-1] und  $GFR_2 = 66,7 \%$  [0,09-0,99] auf Heterogenität ergibt:  $I^2 = 0 \%$  [0-0 %] bzw.  $Q = 0,4795$  mit  $p = 0,4887$ ; beides deutet auf 'nicht heterogen'. Das Poolen gemäß FE erscheint rechnerisch adäquat (4.2.3.5). Die resultierenden Effektschätzer bezgl. RE und FE sind identisch  $GFR_{\text{gepoolt}} = 75,899 \%$ . Das 95 %-CI für  $GFR_{\text{gepoolt}}$  bzgl. RE ist [43,9 %-94,8 %] bzgl. FE [43,7 %-94,9 %]. Beide Intervalle unterscheiden sich praktisch nicht.

**Kritik:** Porcine UBM Scaffolds erweisen sich in der Humanmedizin als Biogerüst hilfreich, etwa zur Unter-

stützung der Regeneration von Weichteilgeweben (z. B. nach chirurgischen Hautverschlüssen, Verbrennungen) oder zur Stimulierung der Fasergewebeproduktion (Sehnen- und Bänderläsionen, Harninkontinenz) sowie zur Förderung des Kieferknochenaufbaus nach Zahnextraktion. Dagegen erzeugt porcine UBM als Pulver-Suspension (ACell™ 0,2-0,4 g Pulver pro 6-19 ml Kochsalzlösung) intraläsional appliziert zwar einen stark angiogenetischen Effekt (LUTZ 2011), kann jedoch eine Immunreaktion mit einer lokal entzündlichen Gewebeschwellung provozieren, was im Vergleich mit anderen prT negativ ins Gewicht fällt (GEBUREK und STADLER 2011b).

#### 5.4.2.2.2 Bei 7 der 11 Paare sind die beiden BehG paar-intern nicht poolbar

**Überblick:** Bei den übrigen 7 (von 11) Paaren 3P, 4P, 5P, 6P, 9P, 10P, 11P erscheint das Poolen der Daten nicht sinnvoll. Die Paare 3P, 4P, 5P, 6P sind zwar bei Gewebestruktur und prT paarweise identisch, jedoch unterscheiden sie sich erheblich bei Prüfzeit, Läsionsgrad oder Altersklasse. In 9P und 10P fehlt die genaue Lokalisation des Gewebeabschnitts des FT. In 11P (4A, 5A) werden in jeder BehG OBS und FT kombiniert. Unter den BehG der 7 nicht poolbaren Paare sind 3 BehG (der 6 BehG unter den 11 Paaren) mit der Fallzahl  $30 \leq N-A$  aus Studie 4A, 5A, Nr. 24.

(1) **Unterschiedliche Prüfzeiten.** Diese Situation liegt bei 4 (von 7) Paaren 3P, 4P, 5P, 6P vor. Bei 3P, 5P, 6P unterscheidet sich die Prüfzeit 12 vs. 24 M paar-intern zwischen beiden BehG erheblich. Bei Paar 4P unterscheiden sich die beiden BehG in der Prüfzeit 6-24 M vs. 'länger als 17' M erheblich und zusätzlich im Läsionsgrad LG 3-4 vs. LG 4-5.

(2) **Fehlende Unterscheidung der Gewebestruktur beim Fesselträger.** Dies trifft auf 2 (der 7) Paare, 9P und 10P zu.

**Paar 9P:** 'FT ohne Angabe der Lokalisation', PRP. Die erste BehG Nr. 7 (ABELLANET et al. 2009) hat  $N = N-A = 19$  Sportpferde (D2, Trab9, S5, V3). Eine Altersangabe der Pferde fehlt. Der Läsionsgrad LG 3-6 streut erheblich von minimal bis schwer. Die zweite BehG Nr. 24 (LUTZ 2011) hat  $N = 71$  Sportpferde (D28, Trab9, S19, V15) mit akuten oder chronischen, minimalen bis leichten Verletzungen LG 3-4\*. Sie ist rein additiv aus drei (1)-(3) BehG derselben Studie Nr. 24 zusammengesetzt: (1) Nr. 24 ( $N = 14$ ) aus Paar 4P am FT-Ursprung; (2) die alleinstehende BehG Nr. 24 ( $N = 15$ ) am FT-Körper; (3) Nr. 24 ( $N = 42$ ) aus Paar 7 am FT-Schenkel. Das Ergebnis  $GFR = 64 \% (45 / 70)$  ist daher undifferenziert und nicht vergleichbar. Die Verteilung der Häufigkeit der Nutzungsrichtung (D, Trab, S, V) unterscheidet sich zwischen beiden BehG signifikant ( $\chi^2$ ;  $p = 0,00465$ ).

**Paar 10P:** 'FT ohne Angabe der Lokalisation', Knochenmarkaspirat. Die Gewebestruktur wird nicht differenziert. Die Prüfzeit 6-18 M, 6 M ist für eine valide Bewertung der Rezidivrate zu kurz. Bei der ersten BehG ist trotz der längeren Prüfzeit 6-18 M bemerkenswerterweise  $GFR_1 = 83 \%$  größer als bei der zweiten BehG  $GFR_2 = 71 \%$  mit der kürzeren Prüfzeit 6 M. Weitere Vergleiche erscheinen nicht sinnvoll, auch da bei der zweiten Gruppe Angaben zum Alter der Pferde und zum Läsionsgrad fehlen.

(3) **Fehlende Unterscheidung der Gewebestruktur OBS und FT.** Dies trifft auf 1 Paar (der 7 Paare) zu, **Paar 11P** (4A und 5A): 'OBS-FT gemischt', Ad-MS. Da unter allen 50 BehG mit Angabe der  $GFR$  nur diese beiden BehG, 4A und 5A, die Gewebestruktur OBS und FT gleichzeitig behandeln und die Untersuchungsergebnisse nicht differenzieren, ist keine mit einer der übrigen 48 BehG vergleichbar.

#### Ergebnis der Heterogenitätsanalyse für 11 Paare zu PrTB

Bei dem Versuch die Daten von BehG zu poolen, bei denen Gewebestruktur und prT identisch sind, ergeben sich aus insgesamt 22 (der 50) BehG 11 Paare. Größere Einheiten als Paare lassen sich in der gleichen Art nicht finden. Nur bei 4 (von 11) Paaren können beide BehG paar-intern gepoolt werden; das entspricht 8

poolbaren (der 22) BehG. Jedoch scheidet 1 (der 4) Paar mit porciner UBM wegen unerwünschter Nebenwirkungen aus. Somit entstehen insgesamt 3 große Beh, die durch die Fallzahl  $30 \leq N-A \leq 145$  potenziell aussagekräftige Ergebnisse für *GFR* haben. 2 (der 3) Paare betreffen die OBS, 1 Paar den FT-Schenkel. Bei den 3 Paaren kommen 2 prT zum Einsatz, nämlich Knochenmark-MSK und PRP. *GFR* = 68,8 % für Knochenmark-MSK bei der OBS Paar 1 ist durch die höchste Fallzahl  $N-A = 145$  Sportpferde am Besten belegt (Tab 5.4.2.2). Die folgende Tabelle Tab 5.4.2.2 fasst das Ergebnis der Heterogenitätsanalyse für 11 Paare zu PrTB zusammen.

Tabelle 5.4.2.2.2 <b>Ergebnis der Heterogenitätsanalyse zu PrTB:</b> Bei 3 (der 11) Paaren können beide BehG gepoolt werden $42 \leq N-A \leq 145$			
	Paar 1	Paar 2	Paar 7
	Knochenmark-MSK	PRP	PRP
	OBS	OBS	FT-Schenkel
<i>GFR</i>	68,8 % [0,61-0,76]	60,4 % [0,47-0,73]	63,7 % [0,48-0,78]
N-A	145	51	42
Verletzung	LG 4-6	LG 3-6	LG 4-5
Nutzung	HR 72 % (105 /145) u.a. (FL 8 HR 105) (D8 S9 FL7 E5 Trab3)	Trab 66 % (35 /51) u.a. (Trab35 S7 R2) (D5 R4 E1)	D17 S11 Trab6 V8
Alter	2-16 a	4-20 a Altersangabe nur 1 BehG	3-24 a
Prüfzeit	24 M; 24 M	12-42 M; 24 M	beide BehG 17 M < P
Nutzungsart: HR Hindernis-Rennen; D Dressur; S Springen; Trab Traber; V Verschiedene. LG Läsionsgrad; a Jahre; M Monate; P Prüfzeit			

#### 5.4.3 Heterogenitätsanalyse zur Behandlungsform KtrB

Die Bedeutung der Behandlungsform KtrB liegt darin, dass sie als positive Kontrolle für die Behandlungsform PrTB geeignet erscheint. Bereits DYSON (2004) vergleicht KtrB u. a. mit einer Hyaluronsäurebehandlung. Allerdings ist es bei der Durchführung klinischer Studien zu equinen Sehnenerkrankungen schwierig für KtrB eine möglichst große Fallzahl zu rekrutieren; oder auch Patienten randomisiert der KtrB-Kontrollgruppe zuzuordnen. Da Patientenbesitzer überwiegend selbst über die Zuordnung entscheiden oder PrTB vorziehen (SMITH 2008; GODWIN et al. 2012).

Zur Behandlungsform KtrB liegen im Zeitraum 2001-2016 nur 4 BehG vor: 3 BehG (1A; Nr. 2; Nr. 3) beziehen sich auf die OBS, *GFR* ist ermittelbar; 2 davon (1A, DYSON 2004; Nr. 3) umfassen Sportpferde unterschiedlicher Nutzungsarten, die dritte (Nr. 2) speziell Cutting Horses. Die vierte BehG bezieht sich auf die Insertion des FT-Schenkels am proximalen Gleichbein bei Thoroughbred Jährlingen (Nr. 28, PLEVIN und MCLELLAN 2014). 2 (der 4) BehG (Nr. 28; 1A) haben die Fallzahl  $46 \leq N-A$ ; *GFR* ist nicht ermittelbar.

Tabelle 5.4.3 **4 alleinstehende BehG der Behandlungsform KtrB**

Legende: [M] Monate, [a] Jahre, LG Läsionsgrad, v vorne, h hinten

Studien-Nr. /Nutzung /Fallzahl	<i>GFR</i>	95 %-CI	Prüfz. [M]	Alter [a]	LG	v /h	Studie
<b>OBS</b>							
$G_{\text{gepoolt: 1A+Nr.2+Nr.3}} = 57,5\% \quad [0,44-0,70]$							
<b>1A</b> FL2 NH 13 E23 S4 End2 D2 24 /46	52 %	[0,37-0,67]	24	ca. 4 -18	4-5*	v46	DYSON (2004)
Nr. 2 Cutting Horses 10 6 /10	60 %	[0,26-0,88]	6-36	3-6	4-5*	k. A.	TIPTON et al. (2013)
Nr. 3 FL1 E1 S1 V2 4 /4	100 %	[0,48-1,00]	24	4-18	3x4*; 6	v5	GEBUREK et al. (2015)

Tabelle 5.4.3 Fortsetzung

Studien-Nr. /Nutzung /Fallzahl	<i>GFR</i>	95 %-CI	Prüfz.	Alter	LG	v/h	Studie
<b>Insertion des FT-Schenkels am proximalen Gleichbein</b>							
<b>Nr. 28</b>	<i>Th 85</i>	<i>k.A.</i> ca. 70 % bei LG 4 bzw. 56 % bei LG 5	6-36	2-3	4-6	v 23 h 35	PLEVIN u. M. (2014)

#### 5.4.3.1 Heterogenitätsanalyse der 4 BehG zu KtrB

1) **Gewebestruktur OBS identisch.** Die Daten der folgenden 3 (der 4) BehG zu KtrB bei der OBS erscheinen poolbar: 1A (DYSON 2004), Nr. 3 (GEBUREK et al. 2015), Nr. 2 (TIPTON et al. 2013).

**Behandlungsgruppe 1A.** *GFR* wird gemäß Modell neu berechnet. Differenziert nach 6 athletischen Nutzungsarten entstehen 6 BehG mit einer kleinen Fallzahl; die *GFR* der 6 BehG streuen maximal zwischen 0 % und 100 %.

Flachrennpferde <i>FL</i>	N-A = 2	A = 0, B = C = 1	<i>GFR</i> = 0 / 2	= 0 % [0,0-0,84]
National Hunt <i>NH</i>	N-A = 13	A = 0, B = 3, C = 5	<i>GFR</i> = 5 / 13	= 38,5 % [0,14-0,68]
Eventing <i>E</i>	N-A = 23	A = B = 0, C = 10	<i>GFR</i> = 13 / 23	= 56,5 % [0,34-0,77]
Distanzreiten <i>End</i>	N-A = 2	A = B = C = 0	<i>GFR</i> = 2 / 2	= 100 % [0,16-1]
Springen <i>S</i>	N-A = 4	A = B = 0, C = 1	<i>GFR</i> = 3 / 4	= 75 % [0,19-0,99]
Dressur <i>D</i>	N-A = 2	A = B = 0, C = 1	<i>GFR</i> = 1 / 2	= 50 % [0,01-0,99]
<u>Gepooltes Ergebnis</u>	N-A = 46	A = 0, B = 4, C = 18	<i>GFR</i> <sub>1A</sub> = 24 / 46	= 52 % [0,38-0,66]

Die statistische Heterogenitätsanalyse der Primärdaten der 6 BehG über alle 6 athletischen Nutzungsarten gepoolt ergibt  $I^2 = 31,7\%$  [0,0-0,72] bzw.  $p = 0,1980$ ; beide Werte deuten auf 'nicht heterogene' BehG (4.2.3.5). Das Poolen gemäß FE erscheint insofern rechnerisch adäquat und ergibt  $GFR_{1A} = 52\%$  [0,38-0,66].

**Poolbarkeit von 1A und Nr. 3.** Die BehG 1A umfasst N-A = 46, Nr. 3 nur N-A = 4 Sportpferde. Die Prüfzeit 24 M ist gleich groß. Läsionsgrad und Altersklasse der Pferde werden in 1A nur angedeutet, können aber mit LG 4-5\* bzw. 4-18\* *a* eingegrenzt werden. In Nr. 3 sind es 'akute' Verletzungen; der Läsionsgrad ist leicht LG 4\* (marginal or diffuse lesions), in einem Fall mittelgradig bis schwer LG 5-6\* (core lesion). Die Autoren von Nr. 3 stellen bezüglich des Alters der Verletzung ein Biasrisiko fest. Gesucht werden zwar Patienten mit ausschließlich akuten Verletzungen, die durch Überbeanspruchung natürlich entstanden sind. Jedoch bildet das Einschlusskriterium 'nicht länger als 14 Tage vor der Erstuntersuchung ist eine entzündliche Gewebeschwellung klinisch angezeigt und zuvor keine therapeutische Injektion verabreicht' die Suche nicht exakt ab. Denn bei Sportpferden zeigen auch subklinische chronische Formen der Verletzung ähnliche Symptome; allerdings nicht in der inflammatorischen Phase, sondern in der frühen Proliferationsphase der Sehnenheilung wie die Autoren kritisch anmerken (GEBUREK et al. 2015). Die Sportpferde in Nr. 3 gehören der jungen bis mittleren Altersklasse (4-18 *a*, Mw = 9,8 *a*) an. Der Wert von  $GFR = 100\%$  erscheint hoch; wahrscheinlich beruht er bei der sehr kleinen Fallzahl 4 Pferde – das 5te Pferd geht während des Follow up verloren – auf einem zufälligen Streuungseffekt, wie er bereits in 1A zu beobachten ist. Zwischen 1A und Nr.3 erscheint die Verteilung der Häufigkeit der Nutzungsrichtung (FL2, NH13, E23, S4, End2, D2 vs. FL1, E1, S1, V2) signifikant unterschiedlich ( $p < 0,001$ ). Was jedoch nicht zwingend gegen das Poolen spricht, da sich die Größe der Fallzahl sehr (46:4) unterscheidet; außerdem stimmen 3 (der 7) Nutzungsarten (FL, E, S) überein. Es ist  $(N-A)_{\text{gepoolt:1A+Nr.3}} = 50$  Sportpferde. Nach dem FE-Modell ist  $GFR_{\text{gepoolt:1A+Nr.3}} = 57,2\%$  [0,43-0,71].

Der Heterogenitätstest ergibt  $I^2 = 80,5 \%$  bzw.  $p = 0,0236$ ; beide Werte deuten auf 'heterogen'. Allerdings ist das  $95 \%$ -CI =  $[0,16-0,95]$  von  $I^2$  sehr breit und somit die Präzision gering. Auch ist Cochran's Q nur schwach sensitiv, da die Anzahl der BehG  $k = 2$  ist. Bei  $k < 10$  ist  $p$  wenig aussagekräftig. Folgt man zunächst dem RE-Modell, so wird das Gewicht der kleinen BehG Nr. 3 von  $9,6 \%$  auf  $42,1 \%$  angehoben; dies führt zu  $GFR_{\text{gepoolt:1A+Nr.3}} = 73,6 \%$ . Das erscheint sachlich nicht adäquat, da beide BehG die Behandlungsform KtrB haben und nicht davon ausgegangen werden kann, dass sich der Einfluss des kontrollierten Bewegungsprogramms zwischen ihnen grundsätzlich unterscheidet. Ein Biasrisiko besteht insoweit, dass die Einhaltung des kontrollierten Bewegungsprogramms in 1 A nicht durchgängig kontrolliert wird.

**Poolbarkeit von 1A, Nr. 3 und zusätzlich Nr. 2.** Ist es sachlich begründet, die Daten von 1A und Nr. 3 zusätzlich mit Nr. 2 zu poolen? Die BehG Nr. 2 umfasst  $N-A = 10$  Sportpferde speziell der Nutzungsart Cutting-Horses; es ist  $GFR = 60 \%$ . Die Prüfzeit ist  $6-36 M$ . Die Pferde gehören der jungen Altersklasse  $3-6 a$  an. Die Verletzungen sind an der OBS peripher lokalisiert und überwiegend leichten LG 4, vereinzelt mittleren Grades LG 5\*, was für die Nutzungsart typisch ist (TIPTON et al. 2013). Da sich unter den  $N-A = 4$  Fällen in Nr. 3 nur ein achtjähriges Pferd mit einer schweren Verletzung 'core lesion' LG 6 befindet, haben somit Nr. 3 und Nr. 2 überwiegend leichte Verletzungen LG 4. Die breit gestreute Prüfzeit  $6-36 M$  (Nr. 2) vs.  $24 M$  (Nr. 3; 1A) ist allerdings nur eingeschränkt vergleichbar. Da Nr. 3 und 1A zusammen ( $N-A = 50$ ) eine sehr viel größere Fallzahl als Nr. 2 ( $N-A = 10$ ) haben, überwiegt der Einfluss der Prüfzeit  $24 M$ . Nach dem FE-Modell ergibt sich  $GFR_{\text{gepoolt:1A+Nr.3+Nr.2}} = 57,5 \%$   $[0,44-0,7]$  und  $(N-A)_{\text{gepoolt:1A+Nr.3+Nr.2}} = 60$  Sportpferde. Die Untersuchung der 3 BehG auf Heterogenität ergibt  $I^2 = 61,1 \%$  bzw.  $p = 0,0767$ .  $I^2$  deutet auf 'möglicherweise heterogen',  $p$  auf 'nicht heterogen' (4.2.3.5). Das breite  $95 \%$ -CI =  $[0-88,9 \%$ ] für  $I^2$  deutet auf ein wenig präzises Ergebnis für  $I^2$ . Die Anzahl  $k = 3$  der BehG ist  $k < 10$ . Deshalb ist der Q-Test wenig sensitiv und hat  $p$  nur eine schwache Aussagekraft. Folgt man dem RE-Modell, werden die Gewichte der beiden sehr kleinen BehG Nr. 3 von  $7,9 \%$  auf  $21,0 \%$  und Nr. 2 von  $17,5 \%$  auf  $31,8 \%$  angehoben, sodass sich  $GFR_{\text{gepoolt}}$  auf  $65,3 \%$  erhöht (Abbildung 5.4.3.1). Die hohe Gewichtung erscheint nicht adäquat. Jedenfalls sind die Einflussfaktoren Läsionsgrad LG (4-5) und Altersklasse (jung-mittel) vergleichbar. Außerdem erfahren alle Pferde dieselbe Behandlungsform KtrB, so dass sich der Einfluss des Bewegungsprogramms nicht grundsätzlich unterscheidet. Abb. 5.4.3.1 vergleicht die Ergebnisse FE, RE für die 3 BehG mittels Forest-Plot graphisch. Der Funnel-Plot lässt keinen Publikationsbias zwischen den 3 BehG erkennen.

Abb. 5.4.3.1 Forest-Plot und Funnel-Plot zu KtrB bei 3 BehG ( $N-A = 60$  Sportpferde)

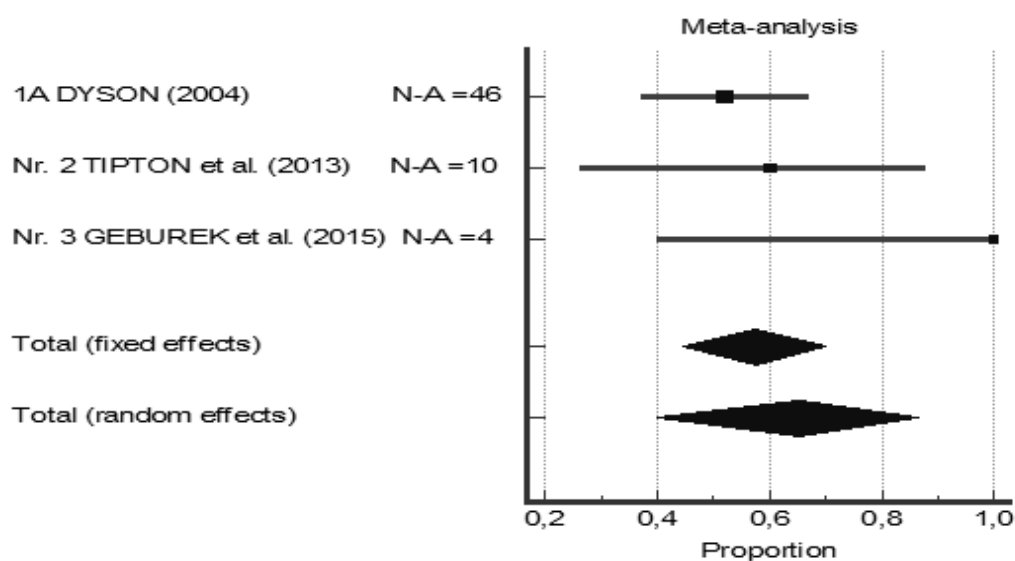
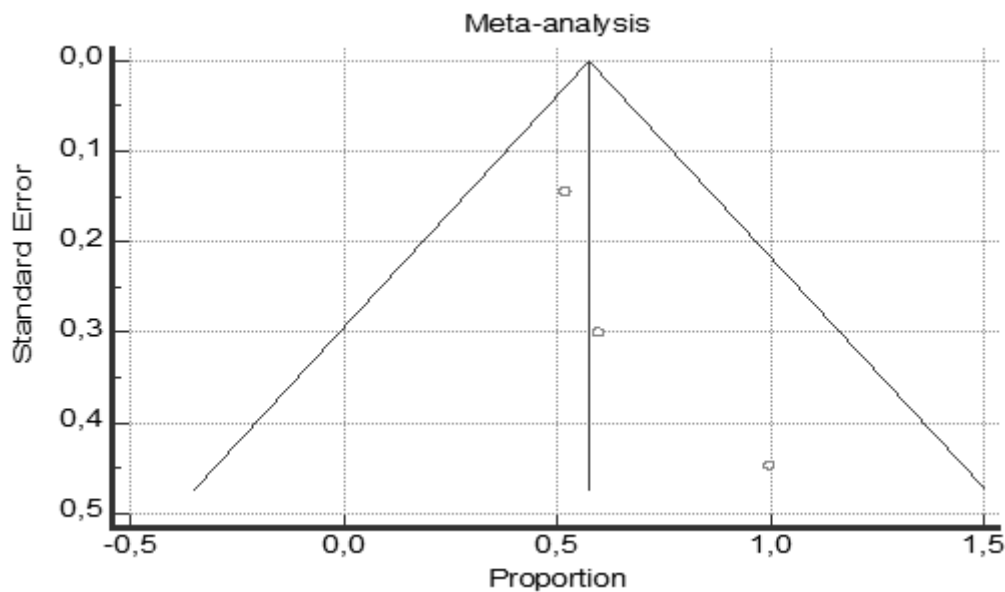




Abb. 5.4.3.1 Fortsetzung Forest-Plot und Funnel-Plot zu KtrB bei 3 BehG (N-A = 60 Sportpferde)



## 2) BehG Nr. 28: Insertion des FT-Schenkels am proximalen Gleichbein

**Überblick:** Studie Nr. 28 (N = 85) untersucht das Erfolgsmerkmal „Ein Pferd nimmt am Ende der dreijährigen Rennsaison noch an Rennen teil“. *GFR* lässt sich nicht ermitteln. Da zwar der Wert der Erfolgsrate  $E'$  sowie der Rezidivrate  $R'$  angegeben wird, jedoch nicht die Bezugsgröße. Das Ergebnis ist daher nicht mit anderen Studien vergleichbar. Jedoch beobachtet die Studie, dass bei abnehmendem Läsionsgrad von LG 5 zu LG 4 die Erfolgsrate  $E_1' = 55,5\%$  auf  $E_2' = 69,6\%$  relativ steigt und die Rezidivrate  $R_1' = 29,2\%$  auf  $R_2' = 10,2\%$  fällt.

**Begründung:** Die retrospektive Kohortenstudie (n = 856) von PLEVIN und MCLELLAN (2014), aus der die BehG Nr. 28 extrahiert ist, misst u. a. die Rennleistung von zwei- und dreijährigen Thoroughbreds (2EPS und 3EPS, average Earnings Per Start). Bei N = 85 Fällen der Kohorte diagnostiziert die Studie beim Training in der Zeit von September des Jährlings-Jahres bis Mai des Folgejahres (ca. 8 M) genau eine Verletzung an der Insertion am proximalen Gleichbein eines FT-Schenkels. Die individuelle Rehabilitationsdauer vor dem ersten Rennen beträgt mindestens 6 M. Die Prüfzeit streut zwischen 10 M und 18 M, da die Studie bis zu Rennen am Ende der dreijährigen Rennsaison dauert. Bei vielen (der 85) Patienten ist trotz kontinuierlicher Ultraschalluntersuchung nicht sicher diagnostizierbar, wann die individuelle Rehabilitationsphase erfolgreich abgeschlossen ist. Das könnte laut Studie die hohen Rezidivraten erklären, da man vermutlich viele Pferde wieder an Rennen teilnehmen lässt, obwohl ihre Läsionen noch nicht vollständig geheilt sind. Am Ende der dreijährigen Rennsaison befinden sich 77,6 % der Kohorte, 69,6 % der Fälle mit LG 4 und 55,5 % der Fälle mit LG 5 noch im Training. Der Anteil der Fälle mit LG 6 wird nicht angegeben, da er laut Studie für ein brauchbares analytisches Ergebnis zu klein ist. Nr. 28 ist unseres Wissens bisher die erste Studie, die zu dieser unter Thoroughbred Jährlingen verbreiteten Verletzungsform Rezidivraten differenziert nach Läsionsgraden vorlegt: Ein Rezidiv vor Beendigung der dreijährigen Rennsaison erleiden 37,5 % der Patienten mit schweren Verletzungen (LG 6), 29,2 % der Patienten mit mittlerem Grad (LG 5) sowie 10,2 % der Patienten mit leichtem Grad (LG 4).

### Ergebnis der Heterogenitätsanalyse für 4 BehG zu KtrB

Bei der OBS können die Daten der 3 (1A, Nr. 3, Nr. 2) BehG zu KtrB gepoolt werden. Da die resultierende Fallzahl  $N-A = 60$  relativ groß ist, hat die gepoolte BehG das potenziell aussagekräftige Ergebnis  $GFR = 57,5 \%$ . Die Tabelle 5.4.3.1 fasst das Ergebnis der Heterogenitätsanalyse aus 4 BehG zu KtrB zusammen.

Tabelle 5.4.3.1 Ergebnis der Heterogenitätsanalyse zu KtrB	
	1 gepoolte BehG (aus 1A, Nr. 2, Nr. 3)
Gewebestruktur	OBS
<i>GFR</i>	57,5 % [0,44-0,70]
N-A	60
Verletzung	LG 4; k. A. bei 1A
Nutzungsart	HR 13 E 23 Cutting Horses 10 u. a.
Alter	3-18 <i>a</i>
Prüfzeit	24; 24; 6-36 <i>M</i>
HR Hindernis-Rennen; E Eventing; LG Läsionsgrad; <i>a</i> Jahre; <i>M</i> Monate	

#### 5.4.3.1.1 Vergleich zwischen Rennpferden und Nicht-Rennpferden

**Rennpferde.** GEBUREK und STADLER (2011b) extrahieren aus Studie 1A die Primärdaten speziell der 15 Rennpferde, nämlich 2 Flachrennpferde mit  $GFR = 0 \%$  und 13 National Hunt Rennpferde mit  $GFR = 38,5 \%$ . Gemäß FE ergeben beide gepoolt  $GFR_{15 \text{ Rennpferde}} = 5/15 = 33,3 \%$  [0,13-0,59].  $I^2 = 25,8 \%$  [0-0 %] deutet auf 'niedrig heterogen' bzw.  $p = 0,2456$  auf 'nicht heterogen' (4.2.3.5). Nimmt man zu 1A noch das einzige Flachrennpferd von Nr. 3  $GFR_{FL} = 100 \%$  hinzu, so ergibt die Heterogenitätsanalyse gemäß FE für  $N-A_{\text{gepoolt}} = 16$  Rennpferde den Wert  $GFR_{16 \text{ Rennpferde}} = 38,7 \%$  [0,18-0,63] (Abb. 5.4.3.1.1a).  $I^2 = 45,1 \%$  [0-0,84] deutet auf 'moderat heterogen' bzw.  $p = 0,1618$  auf 'nicht heterogen' (4.2.3.5). 1A und Nr. 3 stimmen in Prüfzeit 24 *M*, Altersklasse (jung-mittel) und weitgehend auch Läsionsgrad (LG 4-5\*) überein. Abb. 5.4.3.1.1a stellt die metaanalytischen Ergebnisse für die 3 BehG mit insgesamt 16 Rennpferden graphisch mittels Forest-Plot dar. Der Funnel-Plot lässt keinen Publikationsbias zwischen den 3 BehG erkennen.

Abb. 5.4.3.1.1a Forest-Plot und Funnel-Plot zu KtrB bei 3 BehG (N-A = 16 Rennpferde)

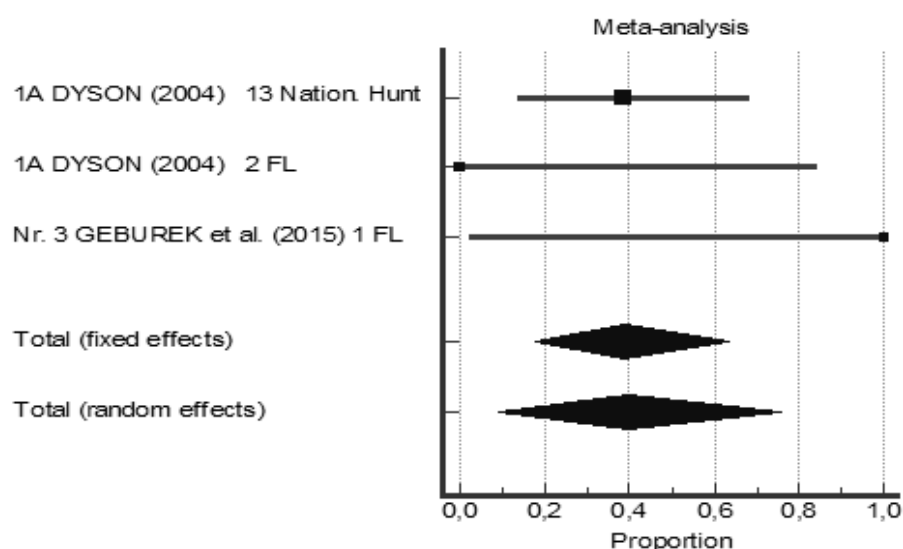
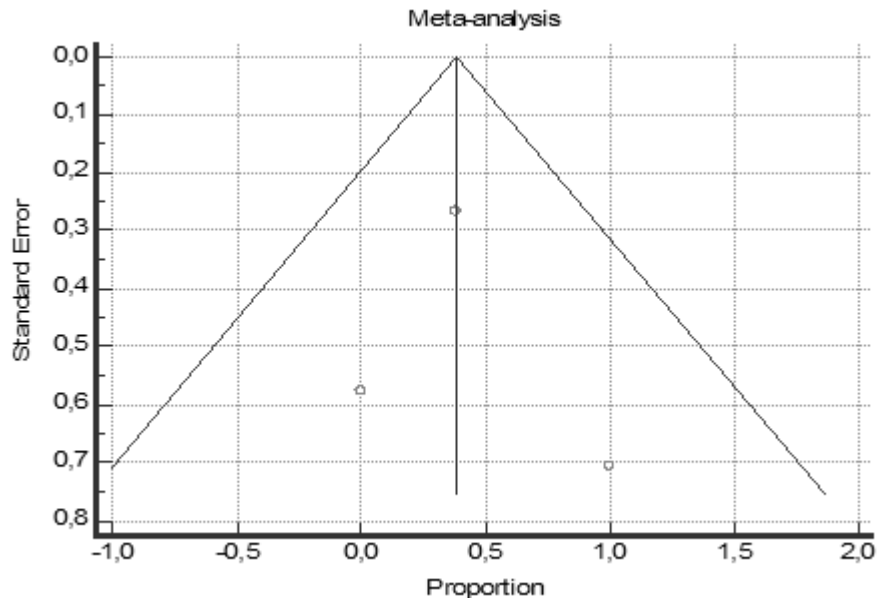


Abb. 5.4.3.1.1a Fortsetzung: Forest-Plot und Funnel-Plot zu KtrB bei 3 BehG (N-A = 16 Rennpferde)



**Nicht-Rennpferde.** Die 44 Nicht-Rennpferde der 3 Studien (1A, Nr. 2, Nr. 3) zu KtrB bei der OBS setzen sich zusammen aus 10 Cutting Horses, 23 E, 2 End, 4 S, 2 D und 3 V\*. Die zugehörigen *GFR*-Werte sind:  $GFR_{10 \text{ Cutting Horses}} = 60 \% [0,26-0,88]$ ;  $GFR_{23 \text{ E}} = 56,5 \% [0,34-0,77]$ ,  $GFR_{2 \text{ DH}} = 100 \% [0,16-1]$ ;  $GFR_{4 \text{ S}} = 75 \% [0,19-0,99]$ ;  $GFR_{2 \text{ D}} = 50 \% [0,02-0,99]$  und  $GFR_{3 \text{ V}^*} = 100 \% [0,29-1,0]$ . Daraus ergibt die Heterogenitätsanalyse gemäß FE für  $N-A_{\text{gepoolt}} = 44$  Nicht-Rennpferde den Wert  $GFR_{44 \text{ Nicht-Rennpferde}} = 64,0 \% [0,49-0,77]$  (Tab 5.4.3.1c).  $I^2 = 0 \% [0-0,74]$  bzw.  $p = 0,4492$  deuten auf 'nicht heterogen'. Da in Nr. 3 von  $N = 5$  Pferden während des Follow up 1 Pferd wegfällt und die Nutzungsart nicht genannt wird, kann unsere Annahme, dass es sich um  $V^* = 3$  Nicht-Rennpferde handelt, einen leichten Bias erzeugen. Wie erwähnt ist die Prüfzeit 6-36 *M* (Nr. 2) vs. 24 *M* (1A und Nr. 3) nur eingeschränkt vergleichbar; jedoch überwiegt durch das Fallzahlverhältnis 10:50 der Einfluss 24 *M*. Die 10 Cutting Horses Nr. 2 sind 3-6 *a* alt, die Pferde Nr. 3 sind ähnlich alt wie 1A 4-18 *a* ( $Mw = 9,8 \text{ a}$ ); ebenso ist der Läsionsgrad (LG 4-5\*) der 3 Gruppen vergleichbar. Abb. 5.4.3.1.1b stellt das Ergebnis der Heterogenitätsanalyse für die 6 BehG mit insgesamt 44 Nicht-Rennpferden graphisch mittels Forest-Plot dar. Der Funnel-Plot lässt keinen Publikationsbias zwischen den 6 BehG erkennen.

Abb. 5.4.3.1.1b Forest-Plot und Funnel-Plot zu KtrB bei 6 BehG (N-A = 44 Nicht-Rennpferde)

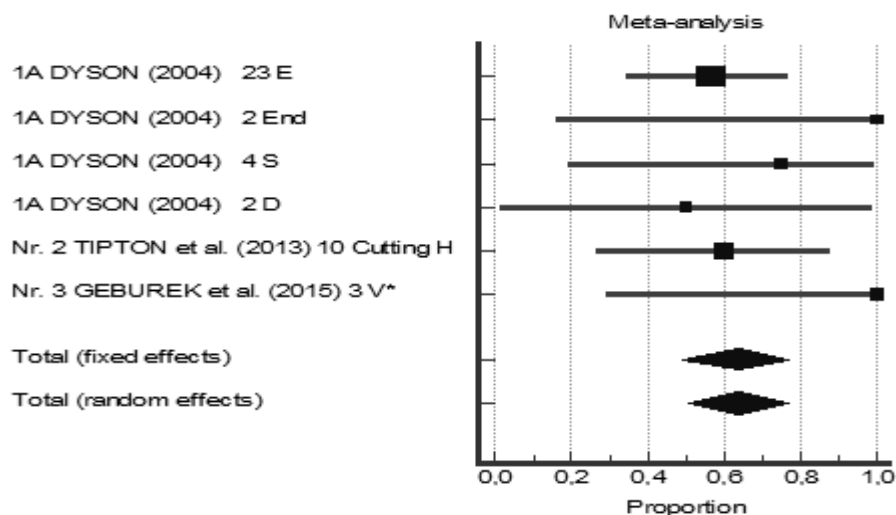
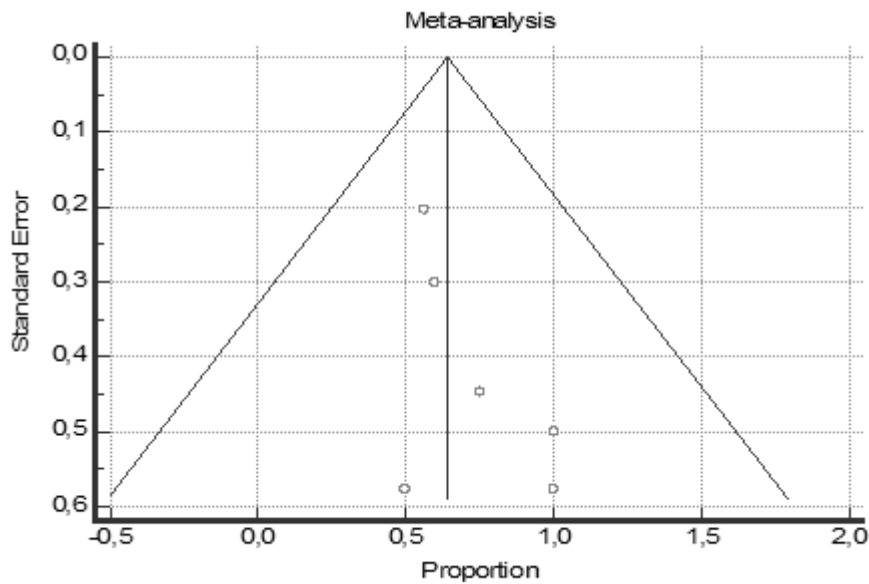


Abb. 5.4.3.1.1b Fortsetzung: Forest-Plot und Funnel-Plot zu KtrB bei 6 BehG (N-A = 44 Nicht-Rennpferde)



Bei der OBS deutet sich zu KtrB für Rennpferde zwar eine kleinere  $GFR = 38,7\%$  an als bei Nicht-Rennpferden  $GFR = 64,0\%$ . Jedoch ist mittels Power-Analyse die Effektdifferenz  $d_{GFR} = 25,3\%$  per Signifikanznachweis durch die gegebene Fallzahl (16 bzw. 44) nicht statistisch gesichert (aktuelle Power  $38,8\%$ ; aktuelles  $p = 0,0366$ ). Andernfalls sind 42 bzw. 115 (total 157) Pferde (aktuelle Power  $80,1\%$ ; aktuelles  $p = 0,0417$ ) oder alternativ 69 Pferde in jeder BehG (total 138 Pferde) erforderlich (aktuelle Power  $80,1\%$ ; aktuelles  $p = 0,0316$ ). Ob mit KtrB Rennpferde und Nicht-Rennpferde unterschiedlich erfolgreich behandelt werden, bleibt daher ungeklärt.

#### 5.4.4 Power-Analyse der Effektdifferenz $d_{GFR}$

Nachdem die Heterogenitätsanalyse der 50 BehG mit insgesamt 11 poolbaren (der 50) BehG abgeschlossen ist, können Unterschiede des Behandlungserfolgs in Form der Effektdifferenz  $d_{GFR}$  zwischen 2 BehG mittels Power-Analyse untersucht und, soweit die erforderlicher Fallzahl gegeben ist, per Signifikanznachweis statistisch gesichert werden. Vergleiche der Fallzahl zwischen den BehG mit Angabe der  $GFR$  zeigen, dass die Power-Analyse die Fallzahlgröße  $30 \leq N-A$  und die Effektdifferenz  $25\% < d_{GFR}$  erfordert. Adäquate paarweise Vergleiche (pV) zwischen 2 BehG setzen voraus, dass qualitative Kriterien erfüllt sind. Idealerweise gehört dazu, dass 2 BehG bis auf den interessierenden Unterschied gleiche Einflussfaktoren auf  $GFR$  haben. BehG, die die Gewebestruktur kombinieren, scheiden grundsätzlich aus.

##### 5.4.4.1 Untersuchung der BehG mit Angabe der $GFR$ auf BehG, die sich möglicherweise für paarweise Vergleiche (pV) eignen

Aus bisherigen Untersuchungsergebnissen der Heterogenitätsanalyse der 50 BehG auf Poolbarkeit folgt:

**Zu PrTB** lassen sich aus 22 BehG 11 Paare bilden, wobei Gewebestruktur und prT paar-intern identisch sind (5.4.2.2). Bei 4 (der 11) BehG-Paaren (nämlich 1P, 2P, 7P, 8P) lassen sich jeweils beide BehG zu einer neuen größeren BehG poolen; jedoch scheidet Paar 8P mit porciner UBM wegen entzündlicher Gewebereaktionen für einen pV aus. Weitere 3 (der 11) BehG-Paare (nämlich 9P, 10P, 11P) kombinieren die Gewebestruktur und scheiden für pV aus. Die 4 verbleibenden (der 11) BehG-Paare (3P, 4P, 5P, 6P) unterscheiden sich paar-intern in der Prüfzeit erheblich. Letztere 8 (= 2 x 4) BehG sind paar-intern nicht poolbar, eignen sich aber möglicherweise für pV mit weiteren (der 50) BehG.

**Zu KtrB** gehören 4 (der 50) BehG; 3 (der 4) BehG lassen sich zu 1 großen BehG bei der OBS poolen; bei der restlichen BehG ist der *GFR*-Wert unsicher (Tab 5.4.3.1). Zu KtrB ergibt sich somit 1 große gepoolte BehG bei der OBS, die sich möglicherweise für pV mit PrTB bei der OBS eignet.

Somit ergeben sich aus 26 (= 22+4 der 50) BehG 12 (= 8+4) BehG, die sich möglicherweise für pV eignen.

Die **restlichen 24** (der 50) BehG gehören alle zu PrTB und sind nicht miteinander poolbar. Sie werden als '24 alleinstehende' BehG bezeichnet und im nächsten Kapitel untersucht. Sie unterscheiden sich durch 7 (der 8) Gewebestruktur-Merkmale (darunter 3 kombinierte Gewebestrukturen) und durch 10 verschiedene prT (darunter 2 prT-Kombinationen). 5 (der 24) BehG kombinieren die Gewebestruktur und scheiden für pV aus. Es verbleiben 19 (der 24) nicht poolbare BehG, die sich möglicherweise für pV eignen.

#### 5.4.4.1.1 '24 alleinstehende' Behandlungsgruppen zu PrTB

Die 24 BehG verteilen sich auf 7 Gewebestruktur-Merkmale wie folgt: OBS 6 BehG, TBS 4 BehG, FT-Ursprung 3 BehG, FT-Körper 4 BehG, FT-Schenkel 2 BehG, FT ohne Angabe der Lokalisation 4 BehG, 'OBS-TBS-USB-FT gemischt' 1 BehG. Es ergeben sich folgende Fallzahlklassen: Bei 14 BehG ist  $N-A < 10$ ; bei 5 BehG ist  $10 < N-A < 15$ ; bei 5 BehG ist  $30 \leq N-A$ . Letztere sind 5 BehG (der 13 BehG unter den 50 BehG) mit der Fallzahl  $30 \leq N-A$ ; sie sind in Tab 5.4.3.4 grafisch (farbiger Hintergrund, fett) hervorgehoben. Somit haben 5 (der 24) BehG die Fallzahl  $30 \leq N-A$ ; 19 (der 24) BehG haben die Fallzahl  $N-A < 30$ .

Tabelle 5.4.4.1.1 '24 alleinstehende' BehG der Behandlungsform PrTB

Legende: [M] Monate, [a] Jahre, LG Läsionsgrad, v vorne, h hinten

Studien-Nr. /prT /Nutzung /Fallzahl	GFR	95 %-CI	Prüfzeit P [M]	Alter [a]	LG	v /h	Studie
<b>OBS</b>							
Nr. 14 Fettgewebe-MSC kombiniert mit PRP <i>S4 Ba2 Rg 1</i>	7 /7	100 % [0,59-1,0]	k. A.	3-15	5-6*	v9	GUERCIO et al. 2015
Nr. 9 Blut-MSC <i>D1 S2</i>	3 /3	100 % [0,29-1,0]	36	10;12;20	5-6*	k. A.	MARFE et al. 2012
Nr. 11 Allog. Amniongewebe-MSC <i>D9 S13 FL6 E4 Trab2</i>	32 /34	94,1 % [0,80-0,99]	24	2-12	5-6	v34	LANGE-C. et al. 2013a
Nr. 13 Knochenmark-MSC kombiniert mit PRP <i>HR 9</i>	6 /9	67 % [0,30-0,93]	24	2-7	4-5*	k. A.	RENZI et al. 2013
Nr. 3 Konditioniertes Blut-Serum <i>G4 D1 V3</i>	5 /8	63 % [0,24-0,91]	24-48	2-20 Mw 8,6	4*	v8	GEBUREK et al. 2015
Nr. 20 Xenogene ovine Amnion-Gewebe-MSC <i>D1 E3 FL5 S3</i>	10 /12	83 % [0,51-0,98]	12	3-17	1-3*	v12	MUTTINI et al. 2013
<b>TBS</b>							
Nr. 11 Allog. Amniongewebe-MSC <i>S2 FL 1</i>	3 /3	100 % [0,29-1,0]	24	2-9	5	v3	LANGE-C. et al. 2013a
Nr. 7 PRP <i>Trab4 Spo2</i>	5 /6	83 % [0,36-1,0]	12-42	k.A.	3-5	k.A.	ABELLANET et al. 2009
Nr. 13 Knochenmark-MSC kombiniert mit PRP <i>HR 5</i>	3 /5	60 % [0,15-0,95]	24	3-8	4-5*	k. A.	RENZI et al. 2013
Nr. 11 Knochenmark-MSC <i>S1 FL 1</i>	1/2-2/2	50-100 % [0,01-1,0]	24	2; 9	5	v2	LANGE-C. et al. 2013a
<b>FT Ursprung</b>							
Nr. 13 Knochenmark-MSC kombiniert mit PRP <i>HR 2</i>	2 /2	100 % [0,16-1,0]	24	2; 5	4-5*	k. A.	RENZI et al. 2013
Nr. 25 Fettgewebe-MSC <i>Spo 30</i>	26 /30	86,6 % < [0,69-0,96]	13-92	1-19 Mw 9	4-5*	k. A.	RICH et al. 2014
Nr. 24 Xenogene UBM <i>D6 S3 Trab1 V4</i>	7 /13	54 % [0,25-0,81]	17 <P	3-24 Mw 10,5	3-4*	k. A.	LUTZ 2011
<b>FT Körper</b>							
Nr. 24 Xenogene UBM <i>D1 S1 Trab5 V1</i>	7 /8	88 % [0,47-1,0]	17 <P	3-24 Mw 10,5	3-4*	k. A.	LUTZ 2011
Nr. 24 PRP <i>D7 S3 Trab2 V3</i>	12 /15	80 % [0,52-0,96]	17 <P	3-24 Mw 10,5	3-4*	k. A.	LUTZ 2011
Nr. 25 Fettgewebe-MSC <i>Spo6</i>	4 /6	67 % [0,22-0,96]	13-92	1-19 Mw 9	4-5*	k. A.	RICH et al. 2014
Nr. 13 Knochenmark-MSC kombiniert mit PRP <i>HR 2</i>	1 /2	50 % [0,01-0,99]	24	4; 8	4-5*	k. A.	RENZI et al. 2013

Studien-Nr.	/prT /Nutzung /Fallzahl	GFR	95%-CI	Prüfzeit P	Alter	LG	v /h	Studie
-------------	-------------------------	-----	--------	------------	-------	----	------	--------

### FT Schenkel

Nr. 25	Fettgewebe-MSC							
	<i>Spo 15</i>	<i>11 /14</i>	<b>79 %</b> [0,49-0,95]	<i>13-92</i>	<i>1-19 Mw 9</i>	<i>4-5*</i>	<i>k. A.</i>	RICH et al. 2014
Nr. 13	Knochenmark-MSC kombiniert mit PRP							
	<i>HR 2</i>	<i>1 /2</i>	<b>50 %</b> [0,01-0,99]	<i>24</i>	<i>5; 6</i>	<i>4-5*</i>	<i>k. A.</i>	RENZI et al. 2013

### Kombination verschiedener Gewebestrukturen

#### FT ohne Angabe der Lokalisation

Nr. 11	Allog. Amniongewebe-MSC							
	<i>D5 Trab1 S8</i>	<i>14 /14</i>	<b>100 %</b> [0,77-1,0]	<i>12</i>	<i>2-12</i>	<i>5-6</i>	<i>v8 h6</i>	LANGE-C. et al. 2013a
Nr. 25	Fettgewebe-MSC							
	<i>Spo 32</i>	<i>29 /31</i>	<b>94 %</b> [0,79-0,99]	<i>13-92</i>	<i>1-19 Mw 9</i>	<i>4-5*</i>	<i>k. A.</i>	RICH et al. 2014
Nr. 24	Xenogene UBM							
	<i>D 11 S5 Trab8 V 10</i>	<i>20 /32</i>	<b>63 %</b> [0,44-0,79]	<i>17 &lt;P</i>	<i>3-24 Mw 10,5</i>	<i>3-4*</i>	<i>k. A.</i>	LUTZ 2011
Nr. 11	Knochenmark-MSC							
	<i>D4 Trab2 S4</i>	<i>3/10-8/10</i>	<b>30-80 %</b> [0,07-0,97]	<i>24</i>	<i>2-12</i>	<i>5-6</i>	<i>v4 h6</i>	LANGE-C. et al. 2013a

#### OBS-TBS-USB-FT gemischt

3A	Knochenmark-MSC							
	<i>V58</i>	<i>43 /58</i>	<b>74 %</b> [0,61-0,85]	<i>12 &lt;P</i>	<i>3-16</i>	<i>k.A.</i>	<i>k.A.</i>	BURK u. BREHM 2011

**Bewertung:** Nur 2 (der 5) BehG mit  $30 \leq N-A$ , Nr. 25 und Nr. 11, unterscheiden die Gewebestruktur, nämlich OBS und FT; die übrigen 3 (der 5) BehG kombinieren die Gewebestruktur und scheiden für pV aus. Die 5 BehG sind: **Nr. 11** OBS mit allogenen Amniongewebe-MSC ( $N-A = 34$ ); **Nr. 25** FT-Ursprung mit Fettgewebe-MSC ( $N-A = 30$ ); Nr. 25 FT ohne Angabe der Lokalisation mit Fettgewebe-MSC ( $N-A = 31$ ); Nr. 24 FT ohne Angabe der Lokalisation mit xenogener porciner UBM ( $N-A = 32$ ); Nr. 3A 'OBS-TBS-USB-FT gemischt' mit Knochenmark-MSC ( $N-A = 58$ ). Von den übrigen 19 (der 24) BehG mit der Fallzahl  $N-A < 30$  kombinieren 2 BehG die Gewebestruktur und scheiden für pV aus.

Somit ergeben sich unter den '24 alleinstehenden' BehG insgesamt 19 BehG, die sich möglicherweise für pV eignen, nämlich 17 BehG mit  $N-A < 30$  und wie erwähnt 2 BehG, **Nr. 11** bzw. **Nr. 25**, mit  $30 \leq N-A$ .

#### Behandlungsgruppe Nr. 25 FT-Ursprung, Fettgewebe-MSC, $N-A = 30$ .

Die Prüfzeit 13-92 *M* (ca. 1-7,7 *a*) streut erheblich. Sie ist viel länger als die Standard-Prüfzeit 18 *M* im *GFR*-Modell für den FT-Ursprung; insofern erscheint der Behandlungserfolg  $GFR = 86,6 \% [0,69-,96]$  hoch. Die Läsionen haben einen leichten bis mittleren Grad LG 4-5\*. Die 30 Sportpferde gehören unterschiedlichen Nutzungsarten an, die in der Studie nicht spezifiziert werden; sie sind aus der jungen bis mittleren Altersklasse (1-19 *a*, *Mw* = 9 *a*). Die Frage studien-interner Biasrisiken wird von den Autoren (RICH et al. 2014) nicht diskutiert.

#### Behandlungsgruppe Nr. 11 OBS, allogene Amniongewebe-MSC, $N-A = 34$ .

Die Prüfzeit beträgt wie im *GFR*-Modell 24 *M*. Der Behandlungserfolg  $GFR = 94,1 \% (32 /34) [0,80-0,99]$  erscheint hoch (2 Rezidivfälle; beide mit LG 6). Die behandelten Läsionen haben einen mittleren bis schweren Grad: 16-mal LG 6 und 18-mal LG 5; sie sind alle an den Vordergliedmaßen lokalisiert. Die 34 Sportpferde gehören unterschiedlichen Nutzungsarten an (D 9, S13, FL6, E4, Trab2). Sie sind aus der jungen und mittleren Altersklasse (2-12 *a*). Die Biasrisiken der Studie (Nr. 11; LANGE-CONSIGLIO et al. 2013a) sind bereits bei Paar 1 genannt, da dort die Vergleichsgruppe aus der Studie Nr. 11 zu Paar 1 gehört; vgl.

#### 5.4.2.2.1.

Tab 5.4.4.1.1a präsentiert die Daten der beiden alleinstehenden BehG Nr. 11 und Nr. 25, die die Gewebestruktur differenzieren und die Fallzahl  $30 \leq N-A \leq 34$  haben.

Tabelle 5.4.4.1.1a <b>2 der '24 alleinstehenden' BehG zu PrTB unterscheiden die Gewebestruktur und haben die Fallzahl <math>30 \leq N-A</math></b>		
	Nr. 11	Nr. 25
	Allogene Amniongewebe-MSC	Fettgewebe-MSC
	OBS	FT-Ursprung
<i>GFR</i>	94,1 % [0,80-0,99]	86,6 % [0,69-0,96]
N-A	34	30
Verletzung	<i>LG</i> 5-6	<i>LG</i> 4-5
Nutzung	D 9 S 13 FL 6 E 4 Trab 2	k. A.
Alter	2-12 <i>a</i>	1-19 <i>a</i> $M_w = 9$
Prüfzeit	24 <i>M</i>	13-92 <i>M</i>
Nutzungsart: FL Flachrennen; D Dressur; E Eventing; S Springen; Trab Traber. <i>LG</i> Läsionsgrad; <i>a</i> Jahre; <i>M</i> Monate		

**Zwischenergebnis** 31 BehG eignen sich möglicherweise für pV zwecks Power-Analyse. Begründung: Die Analyse der '24 alleinstehenden' (der 50) BehG ergibt 19 BehG, die sich möglicherweise für pV eignen. Nimmt man die 12 BehG aus obigem Ergebnis zur Untersuchung der Poolbarkeit der übrigen 26 (der 50) BehG dazu (Erläuterung: 26 BehG = 11 BehG-Paare zu PrTB + 4 BehG zu KtrB), so ergeben sich insgesamt 31 (= 19+12) BehG. Sie bestehen aus 1 gepoolten BehG zu KtrB, 3 jeweils aus einem BehG-Paar gepoolten BehG zu PrTB und 27 (= 19+8) einzelnen BehG zu PrTB. Tabelle 5.4.4.2a stellt im Überblick das Zwischenergebnis der Untersuchung differenziert nach Gewebestruktur, Fallzahlklasse und Poolbarkeit vor.

Tabelle 5.4.4.2a <b>Ergebnis der Untersuchung der BehG mit Angabe der Effektgröße <i>GFR</i> auf BehG, die sich möglicherweise für pV eignen;</b> differenziert nach Behandlungsform (PrTB, KtrB), Gewebestruktur, Fallzahlklasse und Poolbarkeit					
	Anzahl der BehG / Gewebestruktur	N-A < 30	$30 \leq N-A$	Poolbarkeit per Heterogenitäts-Analyse	BehG, die mglw. für pV zwecks <b>Power-Analyse</b> geeignet sind
<b>PrTB</b>	2 FT-Ursprung, 2 FT-Körper, 6 FT-Schenkel, 6 OBS, 6 Kombinationen	16	6	Für 11 BehG-Paare ergibt sich: 4 (der 11 Paare) können zu 4 BehG gepoolt werden; weitere 4 (der 11 Paare) nicht. Weitere 3 (der 11 Paare) kombinieren die Gewebestruktur. Ergebnis: 8 poolbare BehG	3 (der 4) gepoolte BehG mit $42 \leq N-A \leq 145$ . Aus den nicht poolbaren 4 (der 11) BehG-Paare entstehen 8 einzelne BehG mit $N-A < 30$ . Ergebnis: 11 BehG
	3 FT-Ursprung, 4 FT-Körper, 2 FT-Schenkel, 6 OBS, 4 TBS, 5 Kombinationen	19	5	'24 alleinstehende' BehG sind nicht miteinander poolbar. 5 (der 24) kombinieren die Gewebestruktur.	19 (der 24) BehG: 2 (der 19) haben $30 \leq N-A \leq 34$ ; 17 (der 19) haben $N-A < 30$ .
<b>KtrB</b>	3 OBS	2	1	Aus 3 BehG entsteht 1 gepoolte BehG ( $N-A = 60$ )	1 gepoolte BehG $N-A = 60$
	1 FT-Schenkel	0	1	1 nicht poolbare BehG	Vergleichsgruppen fehlen
Summe der BehG		37	13	11 (der 50) BehG	31 BehG



#### 5.4.4.2 Systematische qualitative und quantitative Analyse der 31 BehG auf adäquate pV

Die Analyse, welche der 31 BehG in die engere Wahl für adäquate pV kommen, erfolgt in einem ersten Arbeitsschritt anhand von 2 Fallkonstellationen. Zwischen 2 BehG gilt:

- (1) **Gewebestruktur gleich, aber Behandlungsform (PrTB; KtrB) verschieden;**
- (2) **Gewebestruktur verschieden, aber Behandlungsform (PrTB; KtrB) gleich.**

BehG, die sich weder (1) noch (2) zuordnen lassen, unterscheiden sich in Behandlungsform und Gewebestruktur; pV erscheinen nicht sinnvoll.

In einem zweiten Arbeitsschritt können mittels quantitativer Power-Analyse adäquate pV durchgeführt werden.

Die folgende Zusammenstellung gibt einen Überblick der BehG, die in die engere Wahl für pV kommen, geordnet nach den Merkmalen: Anzahl der BehG, Behandlungsmittel prT, Gewebestruktur, Studiennummer, *GFR* und Fallzahl N-A der regelkonform behandelten Patienten.

Anzahl BehG	prT	Gewebestruktur, Studiennummer, <i>GFR</i> , Fallzahl
6	<b>PRP:</b>	OBS (2P; 60,4 %; N-A = 51); TBS (Nr.7; 83 %; N-A = 6); FT-Körper (Nr. 24; 80 %; N-A = 15); FT-Ursprung (Nr. 23 aus 4P; 80 %; N-A = 20); FT-Ursprung (Nr. 24 aus 4P; 58,3 %; N-A = 12); FT-Schenkel (7P; 63,7 %; N-A = 42).
2	<b>Knochenmark-MSc:</b>	OBS (1P; 68,8 %; N-A = 145); TBS (Nr. 11; 50-100 %; N-A = 2).
5	<b>Knochenmark-MSc kombiniert mit PRP:</b>	OBS (Nr. 13; 67 %; N-A = 9); TBS (Nr. 13; 60 %; N-A = 5); FT-Ursprung (Nr. 13; 100 %; N-A = 2); FT-Körper (Nr. 13; 50 %; N-A = 2); FT-Schenkel (Nr. 13; 50 %; N-A = 2).
3	<b>Fettgewebe-MSc:</b>	FT-Ursprung (Nr. 25; 86,6 %; N-A = 30), FT-Körper (Nr. 25; 67 %; N-A = 6), beim FT-Schenkel (Nr. 25; 79%; N-A = 14).
2	<b>Allogene Amniongewebe-MSc:</b>	OBS (Nr. 11; 94,1 %; N-A = 34), TBS (Nr. 11; 100 %; N-A = 3).
6	<b>Konditioniertes Medium aus allogenem Amniongewebe:</b>	OBS (Nr. 15 aus 3P; 80 %, N-A = 5); OBS (Nr. 16 aus 3P; 80 %; N-A = 5); FT-Körper (Nr. 15 aus 5P; 100 %; N-A = 2); FT-Körper (Nr. 16 aus 5P; 100 %; N-A = 2); FT-Schenkel (Nr. 16 aus 6P; 80 %; N-A = 5); FT-Schenkel (Nr. 15 aus 6P; 75%; N-A = 4).

Die Zusammenstellung ergibt insgesamt 24 BehG mit 6 verschiedenen prT. Verbindet man das tabellarische Ergebnis mit dem quantitativen Kriterium – dass die Power-Analyse die Fallzahl  $30 \leq N-A$  und die Effektdifferenz  $25 \% < d_{GFR}$  erfordert um einen Unterschied des Behandlungserfolgs *GFR* per Signifikanznachweis statistisch sichern zu können – so kommen **Ad (2)** jedoch nur 2 (dieser 24) BehG für 1 adäquaten pV **zu PrTB** in die engere Wahl, nämlich: PRP bei der OBS, 2P vs. PRP beim FT-Schenkel, 7P.

**Ad (1)** Bereits aus den nach der Gewebestruktur geordneten obigen Tabellen (Tab 5.4.2.2, Tab 5.4.3.1, Tab 5.4.4.1.1) ist in Verbindung mit dem quantitativen Kriterium – dass die Power-Analyse die Fallzahl  $30 \leq N-A$  und die Effektdifferenz  $25 \% < d_{GFR}$  erfordert um einen Unterschied des Behandlungserfolgs *GFR* per Signifikanznachweis statistisch sichern zu können – Folgendes zu erkennen: Es kommen nur 4 BehG für adäquate pV in die engere Wahl. Sämtliche 4 BehG beziehen sich auf die OBS, 3 BehG **zu PrTB** und 1 BehG **zu KtrB**: Knochenmark-MSc, 1P, N-A = 145; PRP, 2P, N-A = 51; allogene Amniongewebe-MSc, Nr. 11, N-A

= 34; KtrB, N-A = 60.

**Zwischenergebnis:** Fasst man das Ergebnis zu (1) und (2) zusammen, so kommen 6 BehG, 5 BehG **zu PrTB** und 1 BehG **zu KtrB**, mit der Fallzahl  $30 \leq N-A$  für adäquate pV zwecks Power-Analyse in die engere Wahl.

Die abschließende Prüfung der 6 BehG auf adäquate pV ergibt: Zwischen den 6 BehG sind insgesamt 15 (= 6 über 2) pV möglich. Sie setzen sich wie folgt zusammen. **Ad (1):** Da sich 4 BehG auf die Gewebestruktur OBS beziehen, sind dort 6 (= 4 über 2; der 15) pV möglich. **Ad (2):** 1 weiterer (der 15) pV betrifft PRP bei der OBS und beim FT-Schenkel. **Weder (1) noch (2):** Die übrigen 8 (der 15) pV scheiden aus, da sich Gewebestruktur und prT unterscheiden. Das gilt insbesondere für Nr. 25, Fettgewebe-MSc beim FT-Ursprung (86,6 %; N-A = 30), wofür Vergleichsgruppen mit einer genügend großen Fallzahl fehlen. Die folgende Tabelle (Tab 5.4.4.2b) stellt die Daten der 6 BehG zusammen.

	1A + Nr. 2 + Nr. 3	Paar 1	Paar 2	Nr. 11	Paar 7	Nr. 25
	Kontrollierte Bewegung allein	Knochenmark-MSc	PRP	Allogene Amniongewebe-MSc	PRP	Fettgewebe-MSc
	OBS	OBS	OBS	OBS	FT-Schenkel	FT-Ursprung
<i>GFR</i>	57,5 % [0,44-0,70]	68,8 % [0,61-0,76]	60,4 % [0,47-0,73]	94,1 % [0,80-0,99]	63,7 % [0,48-0,78]	86,6 % [0,69-0,96]
N-A	60	145	51	34	42	30
Verletzung	LG 4 (bei 1A k. A.)	LG 4-6	LG 3-6	LG 5-6	LG 4-5	LG 4-5
Nutzung	HR 13 E 23 Cutting Horses 10 u. a.	HR 72 % (105 /145) u. a.	Trab 69 % (35 /51) u. a.	D 9 S 13 FL 6 E 4 Trab 2	D 17 S 11 Trab 6 u. a.	k. A.
Alter	3-18 a	2-16 a	4-20 a nur 1 BehG	2-12 a	3-24 a	1-19 a Mw = 9
Prüfzeit P	24; 24; 3-36 M	24 M	12-42; 24 M	24 M	17 M < P	13-92 M
Nutzungsart: <b>HR</b> Hindernis-Rennen; <b>D</b> Dressur; <b>E</b> Eventing; <b>FL</b> Flachrennen; <b>S</b> Springen; <b>Trab</b> Traber. <b>LG</b> Läsionsgrad; <b>a</b> Jahre; <b>M</b> Monate; <b>P</b> Prüfzeit.						

#### 5.4.4.3 Ergebnisse der quantitativen Power-Analyse für adäquate pV

Ad (1): Bei 4 BehG ist die Gewebestruktur OBS gleich, aber die Behandlungsform verschieden.

##### **Allogene Amniongewebe-MSC vs. Knochenmark-MSC, vs. PRP, vs. KtrB**

Bei 3 (der 6 bei der OBS möglichen) pV ist zwischen allogenen Amniongewebe-MSC vs. Knochenmark-MSC, vs. PRP sowie vs. KtrB die Effektdifferenz  $25\% < d_{GFR}$  durch die gegebene Fallzahl per Signifikanznachweis statistisch gesichert. Der  $GFR$ -Wert der 3 prT, allogene Amniongewebe-MSC, Knochenmark-MSC, PRP, sowie der  $GFR$ -Wert von KtrB kommen der Zwei-Jahres-Gebrauchsfähigkeitsrate  $2GFR$  des Modells bei der OBS nahe, da die Prüfzeit 24 M übereinstimmt. Wie groß  $2GFR$  tatsächlich ist, muss durch weitere Studien noch genauer ermittelt werden.

Für allogene Amniongewebe-MSC (94,1 %; Nr.11) erscheint vs. Knochenmark-MSC (68,8 %; Paar 1) die Chance  $OR$  für die GF ca. 7-fach größer ( $OR = 7,2 [1,7-31,4]$ ;  $d_{GFR} = 25,3\%$ ; aktuelle Power 94,2 %; aktuelles  $p = 0,0186$ ).

Für allogene Amniongewebe-MSC erscheint vs. PRP (60,4 %; Paar 2) die Chance  $OR$  für die GF ca. 10-fach größer ( $OR = 9,7 [2,0-44,2]$ ;  $d_{GFR} = 33,7\%$ ; aktuelle Power 93,6 %; aktuelles  $p = 0,0190$ ).

Ebenso erscheint für allogene Amniongewebe-MSC vs. KtrB (57,5 %; N-A = 60) die Chance für die GF ca. 11-fach erhöht ( $OR = 11,4 [2,5-52,1]$ ;  $d_{GFR} = 36,6\%$ ; aktuelle Power 97,3 %; aktuelles  $p = 0,0172$ ).

##### **Knochenmark-MSC vs. PRP, vs. KtrB**

Bei den übrigen 3 (der 6 bei der OBS möglichen) pV, nämlich zwischen Knochenmark-MSC vs. KtrB, KtrB vs. PRP und Knochenmark-MSC vs. PRP ist die Effektdifferenz  $d_{GFR} < 25\%$ . Hier ist mangels erforderlicher Fallzahl die kleine Effektdifferenz  $d_{GFR} = 11,3\%$  von Knochenmark-MSC (68,8 %; Paar 1) vs. KtrB (57,5 %) per Signifikanznachweis nicht statistisch gesichert (aktuelle Power 29,6 %; aktuelles  $p = 0,0434$ ). Auch nicht die noch kleinere Effektdifferenz  $d_{GFR} = 2,9\%$  von KtrB vs. PRP (60,4 %; Paar 2) (aktuelle Power  $< 5\%$ ; aktuelles  $p < 0,05$ ). Ebenso nicht die Effektdifferenz  $d_{GFR} = 8,4\%$  zwischen Knochenmark-MSC (68,8 %, 1P) und PRP (60,4 %, 2 P) bei der OBS (aktuelle Power 17,7 %; aktuelles  $p = 0,0447$ ). In letzterem Fall sind in jeder BehG bspw. ca. 530 Sportpferde (total 1060) erforderlich um die kleine Effektdifferenz per Signifikanznachweis statistisch zu sichern (Power 80 %;  $\alpha = 0,05$ ). Ob ein Unterschied zwischen dem Behandlungserfolg mit Knochenmark-MSC, PRP oder KtrB bei der OBS besteht, bleibt daher ungeklärt.

Ad (2): Bei 2 BehG ist PRP gleich, aber die Gewebestruktur, OBS bzw. FT-Schenkel, verschieden. Zwischen PRP (60,4 %, 2P) bei der OBS und PRP (63,7 %, 7P) beim FT-Schenkel ist  $d_{GFR} = 3,3\%$ . Die sehr kleine Effektdifferenz ist bei der gegebenen Fallzahl per Signifikanznachweis nicht statistisch gesichert (aktuelle Power  $< 5\%$ ; aktuelles  $p < 0,05$ ). Andernfalls ist bspw. die Fallzahl ca. 3456 Sportpferde in jeder BehG (total 6912) erforderlich (Power 80,0 %;  $\alpha = 0,05$ ). Ob ein Unterschied zwischen dem Behandlungserfolg mit PRP bei der OBS oder beim FT-Schenkel besteht, bleibt daher ungeklärt.

Die beiden Kreuztabellen Tab 5.4.4.3 zeigen im Überblick das Ergebnis der quantitativen Power-Analyse für die adäquaten 7 (der 15) pV. Bei den übrigen nicht adäquaten 8 (der 15) pV sind Gewebestruktur und prT verschieden; Vergleichsgruppen beim FT fehlen noch. Trotzdem ist bemerkenswert: Der Behandlungseffekt ( $GFR = 86,6\%$ ) mit Fettgewebe-MSC beim FT-Ursprung Nr. 25 (Tab 5.4.4.2b) erscheint ähnlich hoch wie mit allogenen Amniongewebe-MSC (94,1 %) bei der OBS; das wird dadurch noch unterstrichen, dass die Prüfzeit 13-92 M verglichen mit 24 M (Paar 1), 12-24 M (Paar 2), „größer 17“ M (Paar 7) und 24 M (KtrB) bedeutend länger ist.

Tabelle 5.4.4.3		Nr. 11 BehG D 9 S 13 FL 6 E 4 Trab 2 LG 5-6 24 M Prüfzeit		1A, Nr. 2, Nr. 3 gepoolt HR 13 E 23 Cutting-H10 u.a. LG 4, LG-Angabe fehlt bei 1A 24; 12-36; 24 M Prüfzeit
<b>Ergebnisse der Power-Analyse</b>  PrTB vs. KtrB <b>O B S</b>  30 ≤ N-A ≤ 145 GFR = (N-A-B-C) /(N-A) Sportpferde: jung bis mittleres Alter ca. 16 a		Allogene <b>Amniongewebe-MS</b> N-A = 34 2GFR = <b>94,1</b> % [0,80-0,99]		<b>KtrB</b> N-A = 60 2GFR = <b>57,5</b> % [0,44-0,70]
Paar 1 HR 72 % (105 /145) u.a. LG 4-6 24; 24 M Prüfzeit	<b>Knochenmark-MS</b> N-A = 145 2GFR = <b>68,8</b> % [0,61-0,76]	Pw 94,2 %, p = 0,0186 d <sub>GFR</sub> = 25,3 % [9,6-35 %] OR = 7,2 [1,7-31,4]	Pw 29,6 %, p = 0,0434 d <sub>GFR</sub> = 11,3 % OR = 1,6	
Paar 2 Trab 69 % (35 /51) S7 D5 u.a. LG 3-6 12-42; 24 M Prüfzeit	<b>PRP</b> N-A = 51 2GFR = <b>60,4</b> % [0,47-0,73]	Pw 93,6 %, p = 0,0190 d <sub>GFR</sub> = 33,7 % [15,4-47,7 %] OR = 9,7 [2,1-45]	Pw < 5 %, p < 0,05 d <sub>GFR</sub> = 2,9 % OR = 1,1	
1A+Nr. 2+Nr. 3 gepoolt HR 13 E 23 Cutting-H 10 u.a. LG 4, LG-Angabe fehlt bei 1A 24; 12-36; 24 M Prüfzeit	<b>KtrB</b> N-A = 60 2GFR = <b>57,5</b> % [0,44-0,70]	Pw 97,3 %, p = 0,0172 d <sub>GFR</sub> = 36,6 % [17,9-51 %] OR = 11,4 [2,5-52,1]	entfällt	
Bei den hellen Feldern der Kreuztabelle ist die Effektdifferenz 25 % < d <sub>GFR</sub> durch die erforderliche Fallzahl mittels Power-Analyse (90 %; α = 0,025) per Signifikanznachweis statistisch gesichert; <b>Pw</b> aktuelle Power; p aktuelles Signifikanzniveau.				
<b>PRP</b> OBS vs. FT-Schenkel  <b>OBS</b> PRP vs. Knochenmark-MS		Paar 7    FT-SCHENKEL D17 S11 Trab6 V8 LG 4-5 17 M < Prüfzeit	Paar 1    OBS HR 72 % (105 /145) u.a. LG 4-6 24; 24 M Prüfzeit	
		<b>PRP</b> N-A = 42    GFR = <b>63,7</b> % [0,48-0,78]	<b>Knochenmark-MS</b> N-A = 145 2GFR = <b>68,8</b> % [0,61-0,76]	
Paar 2    OBS Trab 69 % (35 /51) S7 D5 u.a. LG 3-6 12-42; 24 M Prüfzeit	<b>PRP</b> N-A = 51 2GFR = <b>60,4</b> % [0,47-0,73]	Pw < 5 %, p < 0,05 d <sub>GFR</sub> = 3,3 % OR = 1,1	Pw 17,7 %, p = 0,0447 d <sub>GFR</sub> = 8,4 % OR = 1,4	
Bei den dunklen Feldern der Kreuztabelle ist d <sub>GFR</sub> < 25 % mangels erforderlicher Fallzahl nicht statistisch gesichert. <b>H</b> horses; Nutzungsart: <b>HR</b> Hindernis-Rennen; <b>D</b> Dressur; <b>S</b> Springen; <b>Trab</b> Traber; <b>V</b> Verschiedene. <b>LG</b> Läsionsgrad; <b>a</b> Jahr; <b>M</b> Monate.				

#### 5.4.5 Zusammenfassung der Ergebnisse der Analyse der 50 BehG mit Angabe der *GFR*

BehG mit *GFR* sind ein wichtiges Teilergebnis der vorliegenden Metaanalyse, da nur mit ihnen die GF der Sportpferde quantitativ valide verglichen werden kann. Aus den 51 Studien in 1983-2016 lassen sich 100 BehG (69 in 2011-2016; 31 in 2001-2011) extrahieren; davon 50 BehG (45 BehG in 2011-2016; 5 BehG in 2001-2011) mit Angabe der Effektgröße *GFR* und Intervallschätzer. Bei Studien in 1983-2001, die alle nur mit KtrB behandeln, kann *GFR* nicht bestimmt werden. 39 (der 50) BehG differenzieren die Gewebestruktur; 11 (der 50) BehG kombinieren verschiedene Gewebestrukturen und sind daher nicht vergleichbar.

Die 50 BehG werden differenziert nach der Behandlungsform PrTB und KtrB auf Poolbarkeit untersucht. Dazu lassen sie sich nach 2 Kategorien ordnen.

Kategorie (A): Aus 22 (der 50) BehG können 11 Paare mit jeweils 2 BehG gebildet werden, so dass Gewebestruktur und prT paar-intern identisch sind. Anhand definierter Poolbarkeitskriterien lassen sich bei 4 (der 11) BehG-Paaren beide BehG paar-intern poolen und mittels quantitativer Heterogenitätsanalyse die resultierenden Effektschätzer  $GFR_{gepoolt}$  incl. Konfidenzintervall ermitteln (Tab 5.4.2.2).

Kategorie (B): Weitere 24 (der 50) BehG unterscheiden sich in Gewebestruktur und prT; sie sind untereinander nicht poolbar und werden als '24 alleinstehende' BehG bezeichnet (Tab 5.4.4.1.1).

Weitere 4 (der 50) BehG gehören zur Behandlungsform KtrB; 3 (der 4) BehG können zu 1 BehG gepoolt werden (Tab 5.4.3.1.b).

Unterschiedliche Auswirkungen eines interessierenden Effekts auf den Behandlungserfolg können bei paarweisen Vergleichen (pV) zwischen 2 adäquaten BehG anhand der Effektdifferenz  $d_{GFR}$  mittels Poweranalyse per Signifikanznachweis statistisch gesichert werden, wenn die erforderliche Fallzahl gegeben ist. 2 BehG sind adäquat, wenn gilt: (1) Die Gewebestruktur ist gleich, aber Behandlungsform (PrTB, KtrB) verschieden; (2) Die Gewebestruktur ist verschieden, aber Behandlungsform (PrTB, KtrB) gleich. Von den insgesamt 50 BehG kommen 31 BehG in die engere Wahl, die die Gewebestruktur differenzieren: 3 aus je 2 BehG gepoolte größere BehG zu PrTB, 1 aus 3 BehG gepoolte größere BehG zu KtrB, 27 einzelne BehG zu PrTB. Da außerdem die Power-Analyse die Fallzahl  $30 \leq N-A$  und die Effektdifferenz  $25 \% < d_{GFR}$  erfordert um einen Unterschied des Behandlungserfolgs *GFR* per Signifikanznachweis statistisch sichern zu können, ergibt die systematische Prüfung der 31 BehG 6 BehG, die zur Beurteilung der GF von Sportpferden per Power-Analyse geeignet sind; nämlich 5 BehG zu prTB und 1 BehG zu KtrB; alle haben die Fallzahl  $30 \leq N-A \leq 145$ . 4 (der 6) beziehen sich auf die OBS, 1 BehG auf den FT-Ursprung und 1 BehG auf die FT-Schenkel (Tab 5.4.4.2b).

Anhand der 6 BehG ergeben sich (von theoretisch 6 über 2 möglichen) 7 adäquate pV. Als Ergebnis der 7 pV (Tab. 5.4.4.3) lässt sich festhalten: Ob sich der Behandlungserfolg mit Knochenmark-MSC, PRP oder KtrB bei der OBS unterscheidet, bleibt ungeklärt, da die erforderliche Fallzahl der Patienten fehlt um die kleine Effektdifferenz  $d_{GFR} < 25 \%$  per Signifikanznachweis statistisch sichern zu können. Aus dem gleichen Grund bleibt ungeklärt, ob sich das regenerative Potenzial von PRP bei OBS und FT-Schenkel bezüglich der GF unterscheidet. Dagegen ist die größere Chance auf einen Behandlungserfolg mit allogenen Amniongewebe-MSC gegenüber 3 Behandlungsformen bei der OBS, nämlich Knochenmark-MSC, PRP, KtrB, wahrscheinlich, da die große Effektdifferenz  $25 \% < d_{GFR}$  per Signifikanznachweis durch die gegebene Fallzahl  $30 \leq N-A \leq 145$  statistisch gesichert ist (aktuelle Power  $> 90 \%$ ; aktuelles  $p \leq 0,025$ ). Das Ergebnis beruht allerdings auf nur einer Studie mit allogenen Amniongewebe-MSC ( $N-A = 34$ ) und bedarf insofern der Bestätigung. Jedoch belegt das Ergebnis exemplarisch die Hypothese, dass sich die GF zwischen PrTB und KtrB sowie zwischen verschiedenen PrTB wahrscheinlich unterscheidet.

Gestützt auf die Ergebnisse der metaanalytische Untersuchung der Studien 1983-2016 in Teil I soll in Teil II das *GFR*-Evaluationmodell systematisch methodisch entworfen werden. Die 'Gebrauchsfähigkeitsrate' *GFR* soll die GF der Sportpferde anhand des Merkmals „Rückkehr zur Arbeit auf dem ursprünglichen Leistungsniveau und rezidivfrei für eineinhalb Jahre beim FT (*1,5GFR*) bzw. für zwei Jahre (*2GFR*) bei OBS, TBS (incl. USB)“ valide quantifizieren. Um zu sichern, dass durch das Modell replizierbare Ergebnisse erzielt werden können, ist es erforderlich (1) das Erfolgsmerkmal (EM) mit den zugehörigen quantitativen Effektgrößen (*E*, *R*, *GFR*) zu definieren, (2) deren Berechnungsweise zu normieren, (3) die Qualität der diagnostischen Parameter zu bewerten und (4) durch Standardisierung der Studiendauer und Prüfzeit die zeitlichen Messbedingungen zur Ermittlung der Effektgrößen (*E*, *R*, *GFR*) zu kontrollieren.

### 5.5 Erfolgsmerkmal, Effektgröße und Störfaktoren des Modells

#### 5.5.1 Definition des Erfolgsmerkmals

Der Behandlungserfolg mit KtrB oder PrTB bei natürlich entstandenen Erkrankungen der OBS, TBS und des FT stellt sich in zwei aufeinanderfolgenden Zeitphasen dar: (1) Die individuelle Rekonvaleszens und Rehabilitationsphase sowie (2) die Prüfphase zur nachhaltigen Kontrolle der erfolgreichen Rehabilitation. Das Erfolgsmerkmal (EM1) der Rehabilitationsphase lautet: 'Der Patient kehrt nach erfolgreicher Rehabilitation zur Arbeit auf dem ursprünglichen Leistungsniveau zurück'; das Erfolgsmerkmal (EM2) der Prüfphase lautet: 'Der Patient bleibt auf dem ursprünglichen Leistungsniveau in der standardisierten Prüfzeit rezidivfrei'. EM1 beschreibt die vorläufige Wirkung, EM2 die nachhaltige Wirkung des Behandlungserfolgs. Jede Komponente des Erfolgs wird durch eine eigene quantitative Effektgröße gemessen. EM1 durch die Erfolgsrate *E*. EM2 durch  $1-R$ .  $1-R$  ist das Komplement der Rezidivrate *R*, welches *R* zum Wert 100 % ergänzt und auch als 'Rezidivfreiheits'-Rate bezeichnet werden kann. *E* beschreibt plausiblerweise den Anteil der regelkonform behandelten Studienpopulation, der nach erfolgreicher Rehabilitation zur Arbeit zurückkehrt;  $1-R$  beschreibt adäquaterweise den prozentualen Anteil davon, der in der Prüfzeit rezidivfrei bleibt.

#### 5.5.2 Definition der 'Gebrauchsfähigkeitsrate' *GFR* aus *E* und *R*

Die Behandlungserfolge zweier prT, prT<sub>1</sub> und prT<sub>2</sub>, sind bei idealerweise gleichen Studienbedingungen durch die bloße Angabe der beiden Erfolgsraten *E*<sub>1</sub> bzw. *E*<sub>2</sub> sowie Rezidivraten *R*<sub>1</sub> bzw. *R*<sub>2</sub> nicht direkt vergleichbar. Beispielsweise impliziert eine niedrigere Rezidivrate *R*<sub>1</sub> (< *R*<sub>2</sub>) allein nicht, dass der Behandlungserfolg mit prT<sub>1</sub> dem mit prT<sub>2</sub> überlegen ist. Denn eine größere Erfolgsrate *E*<sub>2</sub> (> *E*<sub>1</sub>) könnte den Effekt der höheren Rezidivrate *R*<sub>2</sub> ausgleichen; beide prT erzielten möglicherweise denselben Behandlungserfolg. Ein *Fallbeispiel* kann das verdeutlichen. Ist *E*<sub>1</sub> = 80 % und *R*<sub>1</sub> = 25 %, so ist das Produkt  $E_1(1-R_1) = 0,8(1-0,25) = 0,8 \cdot 0,75 = 0,6$ . Ist dagegen *E*<sub>2</sub> = 90 % und *R*<sub>2</sub> = 33,3 %, so ist das Produkt ebenfalls  $E_2(1-R_2) = 0,9(1-0,333) = 0,6$ . Die Vorgehensweise eröffnet in einem ersten Arbeitsschritt die Möglichkeit beide Anteilssraten *E* und *R* zu einem einzigen quantitativen Parameter  $GFR = E(1-R)$  zusammenzufassen. Das Produkt  $E(1-R)$  beschreibt den Anteil der Studienpopulation, der die vorläufige und die nachhaltige Wirkung des Behandlungserfolgs aufweist, somit beide EM (5.5.1) erfüllt. Dieser Anteil der regelkonform behandelten Patienten einer Studienpopulation bleibt über die erfolgreiche Rehabilitation hinaus bis zum Ende der Prüfzeit gebrauchsfähig. Wir schlagen vor das Produkt als 'Gebrauchs-fähigkeits-rate' *GFR* zu bezeichnen.

**Erste Definition von *GFR*:** Das Produkt aus der Erfolgsrate  $E$  gemäß EM1 und dem Komplement der Rezidivrate  $(1-R)$ , der 'Rezidivfreiheits'-Rate gemäß EM2, ergibt die 'Gebrauchsfähigkeitsrate'  $GFR = E (1-R)$ . *GFR* beschreibt den Anteil der regelkonform behandelten Studienpopulation, der nach erfolgreicher Rehabilitation das Erfolgsmerkmal 'Arbeit wieder auf mindestens dem ursprünglichen Leistungsniveau' erreicht und während der Prüfzeit rezidivfrei und damit gebrauchsfähig bleibt.

### 5.5.3 Normierung der quantitativen Effektgröße *GFR*

Natürlicherweise werden (wie erwähnt)  $E$  und  $R$  in zwei aufeinanderfolgenden (getrennten) Zeitphasen gemessen: (1) Individuelle Rehabilitationsphase sowie (2) Prüfphase zur Kontrolle von  $E$  auf ein Rezidiv. Scheiden Studienteilnehmer aus, so ändern sich Größe und Zusammensetzung der Studienpopulation. Damit  $E$  und  $R$  unabhängig davon zwischen zwei BehG vergleichbar sind, wird ihre Berechnung durch drei definierte natürliche Variablen A, B, C normiert. Sie repräsentieren disjunkte Fallgruppen, so dass jeder Patient genau einer Gruppe zugeordnet ist. Ihre Größe kann am Ende der standardisierten Studiendauer leicht abgezählt werden. Die Variablenwerte A, B, C gehören zu den Primärdaten und sollten in einer Studie vollständig präsentiert werden, um die Ergebnisse für *GFR* einschließlich Variabilität valide bewerten zu können. Im Rahmen einer 'gepoolten Re-Analyse' können die Ergebnisse von Studien damit neu bewertet werden. Im Folgenden werden die Variablen A, B, C (5.5.3.1-2) definiert, ihre Eigenschaften beschrieben und die Berechnung von *GFR* daraus hergeleitet (5.5.3.3).

#### 5.5.3.1 Definition der quantitativen Variablen A, B, C

Gemäß *GFR*-Modell werden in einer Studie zu Beginn  $N$  Patienten mit KtrB oder PrTB behandelt. Aus Gründen, die mit dem Behandlungsmittel kausal nicht zusammenhängen, scheiden Patienten aus der Studienpopulation entweder physisch aus (z. B. durch Tod, Zuchtzwecke, Besitzer-Wille) oder werden bei der Berechnung von *GFR* nicht mitgezählt, weil sie nicht regelkonform behandelt werden und durch die Ungleichbehandlung einen Bias erzeugen; sie bilden Gruppe A. Die Anzahl der regelkonform behandelten Patienten ist somit  $N-A$ . Die Patienten B erreichen allein wegen der mangelnden Wirkung des Behandlungsmittels das Erfolgsmerkmal (EM1) 'Arbeit wieder auf mindestens dem ursprünglichen Leistungsniveau' bis zum Ende der standardisierten Studiendauer nicht (z. B. bleiben sie rekonvaleszent).  $N-A-B$  erreichen nach Rehabilitation EM1 und werden danach in der standardisierten Prüfzeit auf ein Rezidiv kontrolliert. Davon erleiden C ein Rezidiv, erreichen somit EM2 ('Der Patient bleibt auf dem ursprünglichen Leistungsniveau in der standardisierten Prüfzeit rezidivfrei') nicht. Insgesamt erreichen somit absolut gezählt  $N-A-B-C$  Patienten sowohl EM1 als auch EM2.

#### 5.5.3.2 Beschreibung der disjunkten Fallgruppen A, B, C

a) **Fallgruppe A.** Laut Definition gehören zu A: (1) Patienten, die das kontrollierte Bewegungsaufbautraining nicht einhalten und somit einen Bias bei der Berechnung von *GFR* erzeugen (GUERCIO et al. 2015). (2) Patienten, die EM1 erreichen, aber nicht die gesamte Prüfzeit auf ein Rezidiv kontrolliert werden können. Weil sie (a) vorher aus der Studienpopulation ausscheiden; oder (b) sich ihre individuelle Rehabilitationsphase zu lange hinzieht. (3) Patienten, bei denen trotz Einsatz präziser Diagnosetechniken (US-Untersuchung, Röntgen etc.) keine sichere klinische Diagnose gestellt werden kann, ob sie „völlig“ klinisch geheilt aus der Rehabilitationsphase in die Prüfphase entlassen werden können. Z. B. kann es sich bei leichteren Verletzungsgraden im Insertionsbereich des FT-Schenkels am proximalen Sesambein als schwierig erweisen den genauen Zeitpunkt für eine erfolgreiche Rückkehr zur Arbeit sicher zu bestimmen, da sich die Heilung des verletz-

ten Sehngewebe mittels US-Kontrolle nur in relativ geringen Veränderungen der Echogenität zu erkennen gibt (PLEVIN und MCLELLAN (2014)). Oder Patienten werden zu früh aus der Rehabilitationsphase zur Arbeit auf dem ursprünglichen Leistungsniveau entlassen und erleiden ein Rezidiv. (4) Patienten, bei denen während der Behandlung zusätzlich Verletzungen an einer anderen oder an der gleichen Gliedmaße, aber an anderer Stelle auftreten. RICH et al. (2014) beobachten, dass sich dadurch die Besserung verzögert und die Wahrscheinlichkeit steigt, dass sich das klinische Ergebnis des Behandlungserfolgs für die gesamte Studienpopulation verschlechtert. Verletzungen insbesondere an der kontralateralen Gliedmaße, die nach der Behandlung auftreten, sind nicht per se ein Fehler der Behandlung (LEPPÄNEN et al. 2009; GODWIN et al. 2012). Sie können mit einer Vielzahl von Faktoren korrelieren. Metabolische Erkrankungen wie Cushing können zur Degeneration der Gewebestruktur des FT führen (HOFBERGER et al. 2015). Auch angeborene oder erworbene Fehlhaltungen oder Spätfolgen eines schwer diagnostizierbaren Belastungstraumas sind mögliche Einflussfaktoren. Würden solche Patienten unter die Zahl der Rezidivfälle C subsumiert (z. B. BURK 2012), führte das zu verzerrten Ergebnissen für *R* und *GFR*. Gegebenenfalls sollten solche Patienten separat in den Studienergebnissen ausgewiesen werden. (5) Patienten, bei denen zwecks retrospektiver Studien interessierende Merkmalsdaten aus Datenbanken rekrutiert, aber die Daten für A, B, C zur Berechnung von *GFR* nur lückenhaft erhoben werden (z. B. LUTTER et al. 2015).

Um die Gefahr eines Attrition Bias (Verlust-Bias) zu vermeiden sollte sichergestellt werden, dass die Patienten der Fallgruppe A nicht durch das gleiche interessierende Merkmal geprägt sind, so dass sich die Zusammensetzung der verbleibenden gegenüber der ursprünglichen Studienpopulation systematisch verändert und den Behandlungserfolg verzerrt. Die Gefahr ist nicht unerheblich, wie das systematische Review von BROSSI et al. (2015) zeigt; es findet unter 11 in vivo Studien (2007-2013) zur PRP-Behandlung von natürlich entstandenen Läsionen bei Sportpferden 1 Studie mit einem ausgeprägten Risiko für Attrition Bias.

b) **Fallgruppe B.** Laut Definition gehören dazu (1) Patienten, die in der Rekonvaleszenz eine Besserung erkennen lassen, dann aber wieder hinter das erreichte Leistungsniveau zurückfallen, so dass sie EM1 (und damit auch EM2) nicht erreichen und rekonvaleszent bleiben oder deren Gesundheitszustand sich nur noch verschlechtert. (2) Patienten, die auf die Behandlung (von Anfang an) nicht ansprechen. In beiden Fällen (1 und 2) verbleiben diese Patienten bis zum Ende der standardisierten Studiendauer (5.7) in der Studienpopulation.

c) **Fallgruppe C.** Solche Patienten erreichen EM1, erleiden dann aber ein Rezidiv. D. h. sie zeigen nach erfolgreicher Rehabilitation und Rückkehr zur Arbeit auf dem ursprünglichen Leistungsniveau die alten Krankheitssymptome (Tendino-, Desmopathie, Lahmheit etc.) an derselben oder einer nahegelegenen Stelle der Gliedmaße, evtl. in verstärkter Form. Die Patienten verbleiben entweder bis zum Ende der standardisierten Studiendauer in der Studienpopulation oder scheiden aus; da sie definitionsgemäß die standardisierte Prüfzeit (18 bzw. 24 M) nicht in voller Länge einhalten müssen (5.7.9). Ein Grenzfall kann das exemplarisch verdeutlichen: Ein Patient erreicht erst unmittelbar vor Ende der standardisierten Studiendauer die Rehabilitation; kurz danach erleidet er in der gerade begonnenen Prüfzeit, aber noch innerhalb der Studiendauer ein Rezidiv und scheidet aus. Zählt man solche Fälle nicht zu C, sondern A, entsteht ein Bias. Denn rechnerisch erscheint bei vergleichsweise gleichem Zähler, aber kleineren Nenner  $GFR = (N-A-B-C) / (N-A)$  größer als es dem tatsächlichen Behandlungserfolg entspricht (5.5.3.3).



Die Berechnungsweise von *E*, *R* und *GFR* wird durch die Variablen *A*, *B*, *C* wie folgt normiert. Die **Erfolgsrate *E*** beschreibt den Anteil der regelkonform behandelten Patienten der Studienpopulation, der nach erfolgreicher Rehabilitation EM1 erreicht. Er wird gemäß  $E = (N-A-B) / (N-A)$  berechnet. Dagegen beschreibt die **Rezidivrate *R***  $R = C / (N-A-B)$  den Anteil der regelkonform behandelten Patienten der Studienpopulation, der EM1 erreicht und danach innerhalb der standardisierten Prüfzeit ein Rezidiv erleidet.

**Zweite Definition von *GFR*:** Aus den Variablen *A*, *B*, *C* kann *GFR* in der Form  $GFR = (N-A-B-C) / (N-A)$  berechnet werden.

Nachweis: Gemäß der ersten Definition ist  $GFR = E(1-R)$ . Ersetzt man *E* und *R* durch die Variablendarstellung, so ergibt sich  $GFR = [(N-A-B) / (N-A)] [(1 - C / (N-A-B))]$ . Das Ausmultiplizieren der eckigen Klammern – unter Beibehalten der runden Klammern – ergibt *GFR* als Differenz zweier Bruchterme:  $GFR = (N-A-B) / (N-A) - (N-A-B) C / ((N-A-B) (N-A))$ . Das Kürzen des zweiten Bruchterms ergibt:  $GFR = (N-A-B) / (N-A) - C / (N-A)$ . Beide Bruchterme haben nun denselben Nenner (*N-A*) und werden zu einem einzigen Bruchterm zusammengefasst. Das ergibt  $GFR = (N-A-B-C) / (N-A)$  q. e. d..

Die zweite Definition hat gegenüber der ersten den Vorteil, dass *GFR* ohne Erfolgsrate *E* und Rezidivrate *R* direkt aus den Variablen *A*, *B*, *C* berechnet werden kann, sofern diese vollständig protokolliert vorliegen.

**Ergebnis:** *GFR* gibt im Rahmen der standardisierten Studiendauer (5.7) an, welcher Anteil der regelkonform behandelten (*N-A*) Patienten sowohl nach erfolgreicher (individueller) Rehabilitation EM1 als auch in der nachfolgenden standardisierten Prüfzeit EM2 erreicht.

#### 5.5.4 Störfaktor freie Zusammensetzung der Studienpopulation

Um einige offenkundige Störfaktoren auf *GFR* zu eliminieren, sollte die Zusammensetzung der Studienpopulation die folgenden 4 Kriterien erfüllen.

(1) Es liegt nahe nur akute, aber keine chronisch-degenerativen Verletzungsfälle in eine Studie gemäß *GFR*-Modell aufzunehmen wie in einigen der 30 hier untersuchten Studien. Um möglichst unverzerrte Ergebnisse zu erzielen sollten Patienten ausgeschlossen werden, bei denen sich Narbengewebe im Bereich der Läsion irreversibel ausgebildet hat, da mit einer kompromittierten katabolischen Mikroumgebung der Zellen zu rechnen ist (VERONESI et al. 2016). Unterschiedlich starke Vernarbungen erzeugen schwer kalkulierbare Störfaktoren auf den Behandlungserfolg (GODWIN et al. 2012). Jedenfalls sollten, wenn beide Arten von Verletzungsfällen untersucht werden, die klinischen Ergebnisse getrennt dokumentiert werden.

(2) In wenigen Studien werden auch Patienten mit einer rezidivierenden Erkrankung untersucht (z. B. DYSON 2004; BURK 2012; ROMAGNOLI et al. 2015). BURK (2012) beobachtet nur bei 50 % solcher Patienten eine erfolgreiche Behandlung. Eine erneute Verletzung der rezidivierenden Gewebestruktur an der alten Stelle oder unmittelbar in der Nähe nach erfolgreicher Rehabilitation kann nicht eindeutig allein der Wirkung des Behandlungsmittels ursächlich zugeschrieben werden. Da bei Läsionen aus einer ursprünglich intakten Sehne verkürzte Sehnenabschnitte mit einer reduzierten Dehnungskapazität entstehen, ist die Gefahr für ein Rezidiv bei gleich starker Belastung wie vor der Verletzung vergleichsweise erhöht. Das *GFR*-Modell schließt solche Patienten aus.

(3) Auch subklinische Formen alterungsbedingter Sehnenerkrankungen sollten ausgeschlossen werden.

(4) VERONESI et al. (2016) betonen, dass die Effizienz einer regenerativen Behandlung nicht nur vom prT, sondern auch vom Heilungspotential des Patienten abhängt, u. a. von Medikamenteneinnahme, systemischen, genetischen oder endokrinen Erkrankungen (z. B. Obesitas, Cushing Syndrom, Östrogenmangel). Dies sollte bei der Zusammensetzung der Studienpopulation berücksichtigt werden.

## 5.6 Qualität der diagnostischen Parameter

Da das *GFR*-Modell evidenzbasiert ist, soll die Qualität der diagnostischen Parameter bewertet werden. Sie bestimmt insbesondere auch das Maß an Sicherheit, mit dem beurteilt werden kann, ob ein Patient aus der Rehabilitationsphase wieder zur Arbeit auf dem ursprünglichen Leistungsniveau entlassen werden kann, ab wann somit die Prüfzeit beginnt. Klinische Untersuchungsergebnisse werden mit quantitativen oder semi-quantitativen Parametern gemessen. Mit semiquantitativen Parametern lassen sich Ergebnisse ähnlich einer Benotungs-Skala in eine geordnete Reihenfolge bringen; ein Ergebnis kann gegenüber einem anderen z. B. als nicht erhöht, leicht erhöht, mäßig oder sehr erhöht (bzw. vermindert) bewertet werden. Ergebnisse unterschiedlicher semiquantitativen Skalen können, da ein absoluter Größenbezug fehlt, nicht oder nur bedingt, d. h. innerhalb einer definierten Umgebung, verglichen werden. Dies birgt das Risiko der subjektiven Einschätzung. Dagegen führen quantitative Parameter mit einem absoluten Größenbezug, wie er bspw. durch den Wert Null auf einer Skala hergestellt wird, zu objektiven vergleichbaren Ergebnissen.

### 5.6.1 Ganganalyse, Entzündungsparameter, Palpationsparameter

Am Bewegungsapparat, OBS, TBS, FT erkrankte Pferde werden bei der Erstvorstellung auf Lahmheitssymptome untersucht. Der Lahmheitsgrad (AAEP American Association of Equine Practitioners, AAEP-Grade 0-5) wird semiquantitativ festgestellt (EDINGER 2010). Die absolute Schwankungsbreite des 95 %-Konfidenzintervalls für die Wiederholbarkeit der Einschätzung des Lahmheitsgrades zwischen Veterinärmediziner beträgt ca. 1,5 AAEP-Grad (KEEGAN et al. 2010). Die Sicherheit einer Einschätzung kann somit nicht beliebig erhöht werden. Sie steigt, wenn (1) Veterinärmediziner dieselbe semiquantitative Skala verwenden und (2) die Fallzahl der Einschätzungen vergrößert wird.

Für geringgradige Lahmheiten, insbesondere an der Hinterhand, ist die subjektive Beurteilung nicht zuverlässig auch wenn sie von erfahrenen Praktikern durchgeführt wird (KEEGAN et al. 2010). Daher wird die Methode der objektiven Ganganalyse empfohlen, welche den Praktiker mittels Inertialsensoren (Beschleunigungssensoren, Kreisel- und Magnetometersensoren) durch quantitative Daten (Beschleunigung der Körperoberfläche, Zeitpunkt und Dauer der Standphase, Stärke und Richtung der Kräfte) bei der Diagnosefindung und Therapieplanung unterstützt (KEEGAN et al. 2010, RUNGSRI et al. 2014, SERRA BRAGANCA et al. 2018). Allerdings geben VAN WEEREN et al. (2017) zu bedenken, dass die kinematische Ganganalyse nur Bewegungssymmetrien und Veränderungen in der Körperposition misst, was nicht notwendig auf Lahmheit und Schmerz hinweist. Dagegen kann die kinetische Untersuchung der Ursache der Bewegung, also der auf den Körper ausgeübten Kräfte, z. B. das Maß der Bodenreaktionskraft (ground reaction force) beim Aufsetzen der Gliedmaße, ein weiterer nützlicher Parameter in der Lahmheitsbeurteilung sein, da die Veränderung des Drucks bei der Lastaufnahme biologisch relevante schmerzhaft Prozesse reflektieren kann (VAN WEEREN et al. 2017). Im evidenzbasierten *GFR*-Modell ist mit letzteren Verfahren der Zeitpunkt sicherer diagnostizierbar, wann ein Patient als 'geheilt' in die Prüfphase entlassen werden kann (5.7).

Entzündungssymptome durch Überwärmung werden mittels Hautoberflächentemperatur quantitativ gemessen. Dagegen werden Entzündungsschmerz, Ödeme und subkutane Schwellungen per Palpation (Dimension, Konsistenz, Elastizität, Mobilität, Druckdolenz) semiquantitativ klinisch bewertet. Meist erfolgt eine Einteilung in die Grade 0-3 'nicht erhöht, leicht erhöht, moderat erhöht, hoch'. Oder wird der Faktor der Dimensionsvergrößerung gegenüber der normalen gesunden Gliedmaße angegeben (SCHMIDT 1989). Semiquantitative klinische Bewertungen sind diagnostisch herausfordernd; z. B. bedeutet das Abklingen der Entzündung in der späten Proliferations- und frühen Remodellierungsphase (ca. Tag 45) nicht notwendig einen Rück-

gang der Schwellung. Auch können nach intraläsionaler Applikation eines prT allein durch den unvermeidlichen Reflux kleinerer Mengen bereits Volumenverringerungen im kutanen, subkutanen und peritendinösen Gewebe beobachtet werden, die aber keine Verringerung des TL-CSA darstellen (GEBUREK et al. 2015).

### 5.6.2 B-Mode-Ultraschalldiagnostik, Messtechnik des UTC-imaging

Mittels sonographischer Erhebung mit der B-Mode-Ultraschalldiagnostik (B-Mode) können Läsionen durch Untersuchungen der Sehnenquerschnitte im Links-rechts-Vergleich bei OBS und TBS diagnostiziert oder im Vorher-nachher-Vergleich subtile Veränderungen im Zusammenhang des klinischen Heilungsverlaufs verfolgt werden. KOJAH et al. (2017) zeigen, dass es einem erfahrenen Untersucher unter Feldbedingungen möglich ist, wiederholt Ultraschallbilder mit großer Treffsicherheit der Lokalität und Übereinstimmung in der Reproduzierbarkeit der Bilderhebung zur Berechnung der Beugesehnenquerschnitte der Vordergliedmaßen zu erstellen und diese mit hoher Präzision richtig auszuwerten.

Allerdings weisen GEBUREK et al. (2016) neuerdings im Rahmen einer RCT (OBS; PRP-Gruppe contra Kochsalz-Gruppe, je Gruppe n = 10; Läsionstypen der PRP-Gruppe: 2 scharf von der Gewebeumgebung abgrenzbare Läsionen, 'core lesions'; 8 marginale bzw. diffuse Läsionen) auf die geringe Sensitivität der B-Mode Messtechnik hin. Sie erschwert es feinstrukturierte Veränderungen während des Heilungsprozesses der ersten Wochen genau zu messen.

Für feinstufige hochauflösende Diagnosezwecke der Gliedmaßensehnen ist daher die Messtechnik des UTC-imaging (4 Echotypen) zu empfehlen. Sie verfügt über Messstufen im Abstand von 0,2 mm über eine Länge von 12 cm. Die 'Ultrasound Tissue Characterization' ist eine computergesteuerte Gewebebeurteilung von Ultraschallbildern mit einer deutlich höheren Sensitivität als B-Mode. Mittels Software wird eine 3D-Farbdarstellung der gescannten Sehne erstellt. Normale Bereiche mit intakten gleichgerichteten Sehnenfibrillen (Faszikel, sekundäre Kollagenbündel) werden grün, defekte Gewebazonen, erzeugt von nicht völlig gleichgerichteten und /oder angeschwollenen Sehnenfasern werden rot und Narbengewebe erzeugt von Sehnenmatrix hauptsächlich bestehend aus Sehnenfibrillen, die nicht oder noch nicht zu Fasern organisiert sind, werden blau dargestellt. Schwarz bleiben Bereiche erzeugt von Sehnenmatrix hauptsächlich mit amorpher Struktur oder Flüssigkeit. Zur genaueren Diagnosestellung kann jede Position vergrößert angesteuert werden. Nach VAN SCHIE et al. (1999) sind die Echogenitäts-Scores und Faserausrichtungs-Scores, die auf B-Mode beruhen, mit den speziellen Echogenitäts-Graden von UTC erwiesenermaßen nicht kompatibel. Die Behandlungskosten bei UTC sind beträchtlich höher. Andererseits eröffnet die Messtechnik bessere diagnostische Möglichkeiten.

### 5.6.3 Läsionsgrad nach RANTANEN et al. (2011)

Wie können die Qualität (Lage, Typ, Grad etc.) von akuten Erkrankungen bei OBS, TBS und FT sowie die Veränderungen während des Heilungsprozesses valide gemessen werden?

Zur Frage liegen seit den 1980er Jahren auf der Basis der klinischen B-Mode-Ultraschalldiagnostik Lösungsvorschläge insbesondere der Arbeitsgruppe um RANTANEN und GENOVESE vor (GENOVESE 1986; RANTANEN et al. 2011: Ultrasonographic Evaluation of the Equine Limb, Technique, Diagnosis and Management of Lameness in the Horse). RANTANEN et al. (2011) evaluieren zwecks Befunderhebung quantitative und semiquantitative Parameter als Einzel- oder Summenparameter. Grundlegende Voraussetzung ist der Links-Rechts-Vergleich mit der gesunden kontralateralen Gliedmaße. Aktuelle Verletzungen lassen sich gemäß 6 Kategorien valide klassifizieren sowie Veränderungen während des Heilungsprozesses bewerten. Tab 5.6.3 präsentiert die Kriterien der 6 Kategorien im Überblick. Inwieweit die Nutzung der Sportpferde wieder fortgesetzt oder das kontrollierte Bewegungsmangement ausgedehnt oder aber die Ruhephase verlängert

werden sollte, kann im Einzelfall hinreichend genau prognostiziert werden. Die Sehnen und Bänder der Gliedmaße werden von proximal nach distal im Bereich metacarpus [metatarsus] bis zum Fesselkopf in sieben [neun] gleichlange (4-5 cm) Zonen unterteilt; es ergibt sich der alpha-numerische Zonen-Score 1A/B, 2A/B, 3 A/B/C [3A/B, 4 A/B/C] (GENOVESE 1986); hinzu kommen im Bereich zwischen Fessel- und Hufgelenk weitere vier Zonen P1A/B/C, P2 (P 'pastern').

#### 5.6.3.1 Semiquantitative Parameter

**ES (echo score).** Die Qualität eines sonographischen Echobefundes hängt neben den Fähigkeiten des Untersuchers (Handhabung des US-Gerätes; Wahl des geeigneten Beschallungswinkels, diagnostische Fähigkeit etc.) u. a. von der technischen Ausstattung (Bildschirm, Schallkopf etc.), der Frequenz (Auflösungsgrad, Beschallungstiefe) und der Beschaffenheit der gescannten Gewebeunterlage ab. ES ist ein ganzzahliges semi-quantitatives Maß für die Echogenität von (wasserhaltigem) Körpergewebe und eignet sich zur Erstbeurteilung der erkrankten Gewebestruktur mit Links-Rechts- und zur Verlaufskontrolle mit Vorher-Nachher-Vergleich. Die Prüfskala hat 4 ES-Grade (GEBUREK et al. 2016; VAN SCHIE et al. 1999). Bezogen auf Seh-  
nengewebe bedeutet ES = 0 iso-echogen oder normal-echogen; ES = 1 geringgradig hypo-echogen; ES = 2 gemischt-echogen, d. h. 50 % echogen und 50 % anechogen; ES = 3 bedeutet überwiegend hypoechogen bis total anechogen. Die ES-Grade aller Zonen einer Gewebestruktur ergeben addiert den Summenparameter **T-ES 'total echo score'**. Steigt der ES, nimmt die Echogenität ab und umgekehrt. Da der ES u. a. aufgrund von zufällig verteilten Impedanzunterschieden, Inhomogenitäten der Gewebestruktur und der antagonistischen Wechselwirkung zwischen Frequenzhöhe, Eindringtiefe und Auflösungsvermögen des Ultraschalls bisher nicht quantitativ standardisierbar erscheint und subjektiv geschätzt werden muss, sind ES-Ergebnisse nur bedingt vergleichbar.

**FAS** (fiber alignment score). Um die Qualität des Sehngewebes in longitudinaler Ausrichtung zu bewerten, wird der prozentuale Anteil der parallel ausgerichteten Kollagenfaserbündel am gesamten Faseraufkommen einer Zone subjektiv geschätzt. Die stetig verteilten Anteile werden nach ihrer Quantilszugehörigkeit (25 %-, 50 %-, 75 %-Quantil) klassifiziert; es entsteht der vierstufige ganzzahlige FAS = 0-3. FAS = 0 bedeutet mindestens 75 %, dagegen FAS = 3 höchstens 25 % Anteil parallel ausgerichteter Fasern. Beispielsweise beobachtet DYSON (2004) bei 68 Sportpferden mit Tendinitis der OBS, dass der FAS der MIZ (maximal injury zone) nach 4 Monaten Rehabilitationsphase ein guter Prädiktor für ein Rezidiv ist. Werden die ganzzahligen FAS-Werte aller Zonen einer Gewebestruktur addiert, entsteht der Summenparameter T-FAS (total fibre alignment score). Teilt man den T-FAS durch die Anzahl der Zonen, so ergibt sich der empirische Mittelwert A-FAS (average fiber alignment score); er ist ein semiquantitatives Maß für den geschätzten durchschnittlichen Anteil der parallel ausgerichteten Fasern mehrerer Zonen einer Gewebestruktur. Zur Verlaufskontrolle lassen sich im Vorher-Nachher-Vergleich derselben Gewebestruktur die Veränderungen im Heilungsprozess mit FAS gut dokumentieren. FAS eignet sich nur bedingt für den Vergleich der Ergebnisse unterschiedlicher Studienpopulationen.

### 5.6.3.2 Quantitative Parameter

**CSA-Differenz im Links-Rechts-Vergleich.** Die CSA-Werte ( $\text{mm}^2$ ; cross-sectional area) der verletzten Gliedmaßenzone sowie der entsprechenden Zone der gesunden kontralateralen Gliedmaße werden verglichen.

**MIZ-CSA.** Die Zone mit dem maximalen Anteil der verletzten Fläche an der Querschnittsfläche heißt MIZ (maximal injury zone); das ist die Zone mit der 'größten' Verletzung, somit der vergleichsweise am stärksten eingeschränkten physiologischen Funktion. Die absolute Größe der Querschnittsfläche der MIZ ist die MIZ-CSA (MIZ cross-sectional area); die MIZ-L-CSA (MIZ lesion cross-sectional area) ist die absolute Größe der verletzten Querschnittsfläche der MIZ. Der Quotient  $(\text{MIZ-L-CSA} / \text{MIZ-CSA}) \times 100 \%$  beschreibt den prozentualen Flächenanteil der MIZ-L-CSA an der MIZ-CSA und ist ein quantitatives Maß, das sich für den objektiven Vergleich der Ergebnisse unterschiedlicher Studienpopulationen eignet; damit lassen sich auch Veränderungen beim Heilungsprozess valide quantitativ beschreiben.

**T-Lesion %.** Addiert man die CSA-Werte bzw. die L-CSA-Werte ( $\text{mm}^2$ ; lesion cross-sectional area) aller verletzten Gliedmaßenzonen einer Gewebestruktur (OBS, TBS, FT etc.), so entstehen die Summenparameter T-CSA (total cross-sectional area) bzw. TL-CSA (total lesion cross-sectional area). Der Quotient  $(\text{TL-CSA} / \text{T-CSA}) \times 100 \%$  ist der T-Lesion % (percent total lesion). Er erlaubt die genaueste Quantifizierung der Gewebeschädigung und eignet sich für einen objektiven Vergleich der Ergebnisse unterschiedlicher Studienpopulationen sowie der Veränderungen beim Heilungsprozess.

### 5.6.3.3 Verletzungskategorien: Läsionsgrade (LG)

Die Stärke der von RANTANEN et al. (2011) entwickelten LG-Skala liegt darin, dass die Einteilung in 6 LG valide auf einem quantitativen Messparameter CAS (cross-sectional area,  $\text{mm}^2$ ) basiert, der verletzten Querschnittsfläche einer Gliedmaßenzone. MIZ-CSA und T-Lesion % werden aus CSA berechnet. Die Einteilung in LG wird durch den semiquantitativen Parameter ES differenzierend unterstützt. Die Gefahr subjektiver Interpretationsfehler ist dadurch systematisch reduziert. Im Links-Rechts-Vergleich können bereits kleine Querschnittsflächenvergrößerungen (LG 1-3) und feinstrukturierte echogene Gewebeanomalien (LG 2-3) durch beide Parameter unabhängig detektiert und die Befunde einer Verletzungs-Kategorie zugeordnet werden. Der semiquantitative Parameter FAS (0-3) beschreibt darüber hinaus den Anteil an parallel ausgerichteten Gewebefasern in longitudinaler Richtung.

**Kategorie 1.** ES und FAS ist Null bei jeder von der Anomalie betroffenen Gewebezone. Im Links-Rechts-Vergleich mit der kontralateralen gesunden Gliedmaße ist  $\text{CSA} > 25 \%$  als klinisch substantielle Vergrößerung einer Gliedmaßenzone definiert. Bei keiner Zone jedoch sollte die Vergrößerung  $\text{CSA} > 39 \%$  sein; dies erlaubt eine gewisse Messfehlertoleranz. Je mehr Zonen betroffen sind, desto geringer ist der T-CSA-Unterschied zwischen den Gliedmaßen; bspw. ist bei 3 Zonen  $\text{T-CSA} < 20 \%$ , bei 6-7 Zonen  $\text{T-CSA} < 15 \%$  (Tab 5.6.3).

**Kategorie 2.** Wie bei Kategorie 1 sind keine hypo- oder anechogenen Faserbündel der erkrankten Sehnen- oder Bandstruktur erkennbar; es ist  $\text{ES} = \text{FAS} = 0$ . Obligates Hauptunterscheidungsmerkmal gegenüber Kategorie 1 ist die CSA-Vergrößerung der MIZ um mehr als 39 %. Fakultatives Unterscheidungsmerkmal gegenüber LG 1 ist: T-CSA kann bei 6-7 Zonen mehr als 14 % betragen, also größer sein als bei LG 1 oder auch nicht. Krankhafte Veränderungen der Kategorie 2 treten klinisch als lokal abgegrenzte oder diffus verteilte Verdickung der Sehne oder des Bandes in Erscheinung (Tab 5.6.3).

**Kategorie 3.** Im Vergleich zu LG 2 ist die klinisch signifikante MIZ-CSA-Vergrößerung oder die T-CSA-Vergrößerung mehr als LG 1 fakultativ. Hauptunterscheidungsmerkmal gegenüber LG 2 sind fokale hypo-echogene Sehnenfaserbündel ES = 1-2 bei einer, oder im Falle mehrerer betroffenen Zonen T-ES < 2. Krankhafte Veränderungen der Kategorie 3 treten im Anfangsstadium einer minimalen Verletzung oder im Endstadium einer abgeheilten Verletzung auf. Wie bei Verletzungen der Kategorie 2 ist die Chance auf vollständige Genesung hoch, wenn rechtzeitig ausreichend Ruhephasen gewährt werden.

**Kategorie 4** umfasst leichte Verletzungen. Die MIZ-CSA kann klinisch signifikant, d. h. um mehr als 39 % vergrößert sein, muss aber nicht. Auch die T-CSA-Vergrößerung größer als LG 1 ist fakultativ. Eines der beiden Hauptunterscheidungsmerkmale gegenüber LG 3 ist: Die Läsionen erscheinen als fokal hypo- oder anechogene Bereiche ES = 3 bei genau einer bzw.  $3 \leq$  T-ES bei mehreren Zonen. Als zweites Unterscheidungsmerkmal kommt der quantitative Parameter T-Lesion % ab LG 4 neu hinzu. Ab 3 verletzten Zonen ist T-Lesion % 1-25 %. Mit wachsender Anzahl der Zonen nimmt der Wert ab; z. B. beträgt er bei 11 verletzten Zonen T-Lesion % 1-10 % (Tab 5.6.3).

**Kategorie 5** umfasst moderate (mäßige) Verletzungen. Die beiden Hauptunterscheidungsmerkmale gegenüber LG 4 sind: (1) Die MIZ-CSA ist substantiell, d. h. um mehr als 39 % vergrößert; auch die T-CSA-Vergrößerung über LG 1 hinaus ist obligat. (2) Ab 3 verletzten Zonen ist T-Lesion % 26-35 %; das ist größer als bei LG 4. Mit wachsender Zahl an Zonen nimmt der Wert ab; z. B. beträgt er bei 11 verletzten Zonen T-Lesion % 11-20 %. In den Kategorien 5-6 lassen viele Pferde das Anfangsstadium von Lahmheit erkennen; in den Kategorien 3-4 dagegen nur wenige.

**Kategorie 6** umfasst schwere Verletzungen. Wie Kategorie 5 sind die MIZ-CSA-Vergrößerung und die T-CSA-Vergrößerung über LG 1 hinaus obligat. Die verletzten Gewebestrukturen erscheinen noch stärker hypo- bis anechogen. Hauptunterscheidungsmerkmal gegenüber LG 5 ist: Ab 3 verletzten Zonen gilt T-Lesion % > 35 %; das ist größer als bei LG 5. Mit wachsender Zahl nimmt der Wert ab; bei z. B. 11 verletzten Zonen gilt T-Lesion % > 20 %.

Tabelle 5.6.3		Läsionsgrad (LG) nach RANTANEN et al. (2011)										
LG		ES				Ab LG 1 CSA-V bzw. MIZ-CSA-V verglichen mit der gesunden kontralateralen Gliedmaße Ab LG 4 Zusätzlich T-Lesion % der verletzten Gliedmaße						
		0 isoech.	1 gering hypoech.	2 50 % ech. 50 % anech.	3 anech.	1 verletzte Zone	2 verletzte Zonen	3 verletzte Zonen	5 verletzte Zonen	6-7 verletzte Zonen	11 verletzte Zonen	
1	mini- mal	FAS = 0	—		—	CSA-V > 25	T-CSA-V < 25	T-CSA-V < 20	T-CSA-V < 17	T-CSA-V < 15	T-CSA-V < 12	
2		FAS =0	—			MIZ-CSA-V > 39 obligat			T-CSA-V > LG 1 fakultativ			
3		—	T-ES < 2			MIZ-CSA-V fakultativ			T-CSA-V > LG 1 fakultativ			
4	leicht	—			3 ≤ T-ES	T-Lesion %		1-25	1-20	1-15	1-10	
						MIZ-CSA-V fakultativ			T-CSA-V > LG 1 fakultativ			
5	mittel	—				T-Lesion %		26-35	21-30	16-25	11-20	
						MIZ-CSA-V obligat			T-CSA-V > LG 1 obligat			
6	schwer	—				T-Lesion %		> 35	> 30	> 25	> 20	
						MIZ-CSA-V obligat			T-CSA-V > LG1 obligat			
CSA cross-sectional area [mm²]; ech. echogen; ES echo-score; FAS fiber alignment score; LG Läsionsgrad; MIZ maximal injured zone; T total; T-Lesion % percent total lesion ab LG 4; V Vergrößerung; Zahlenwerte für T-Lesion % in Prozent.												

## 5.7 Kontrolle der Messbedingungen zur Bestimmung von $E$ und $R$ – Standardisierung der Studiendauer

Um mit dem *GFR*-Modell vergleichbare Ergebnisse erzielen zu können, ist die Standardisierung der Studiendauer erforderlich. Die Studiendauer summiert sich aus der Rehabilitationsdauer und der Prüfzeit. Daher stellt sich die Aufgabe den Zeitbereich der Rehabilitationsdauer durch einen geeigneten validen Schätzwert, die 'projektierte Rehabilitationsdauer' (proj. Reha), sinnvoll einzugrenzen und in einem zweiten Schritt die Prüfzeit zu standardisieren.

Bei der Rehabilitationsdauer können 3 Aspekte unterschieden werden: (1) die 'individuelle Rehabilitationsdauer' des einzelnen Patienten, (2) die 'Rehabilitationsdauer der Studienpopulation' und (3) die 'projektierte Rehabilitationsdauer'.

(1) Die **individuelle Rehabilitationsdauer** (5.7.1-2) ist die Zeitspanne, die der einzelne Patient nach Behandlung mit PrTB oder KtrB bis zur Rückkehr zur Nutzung auf dem ursprünglichen Leistungsniveau benötigt. Die Rehabilitation wird von einem kontrollierten Bewegungsprogramm unterstützt, das an die Nutzungsrichtung und den individuellen Heilungsfortschritt angepasst ist. Diese Zeitspanne ist naturgemäß nicht festlegbar. Jede Festlegung würde dazu führen, dass Patienten, die länger benötigen, bei der Erfolgsermittlung negativ ins Gewicht fallen, dagegen Patienten mit kürzerer Rehabilitationsdauer selektiv bevorzugt werden. Da dies den empirischen Effektschätzer der Erfolgsrate  $E$  bezüglich der Grundgesamtheit der Sportpferde systematisch verzerrt, würde das Modell unbrauchbar. Daher begrenzt das Modell im Rahmen der standardisierten Studiendauer die individuelle Rehabilitationsdauer nicht willkürlich; d. h. diese kann maximal die gesamte Studiendauer beanspruchen.

(2) Die **Rehabilitationsdauer der Studienpopulation** ergibt sich als empirischer Mittelwert (zufälliges arithmetisches Mittel) der individuellen Rehabilitationsdauer der behandelten Patienten einer Studienpopulation. Die vorliegenden Studien geben diesen überwiegend nicht an. Oft wird die Zeitspanne zwischen Minimal- und Maximalwert der individuellen Rehabilitationsdauer berichtet oder nur der Maximalwert.

(3) Die **projektierte Rehabilitationsdauer** (5.7.4) ist ein valider Schätzwert, der sich idealerweise als Erwartungswert der Verteilung der 'Rehabilitationsdauer der Studienpopulation' der 30 relevanten Studien im Zeitraum 2011-2016 ergibt. Dieser ist aufgrund der heterogenen Angaben zur Rehabilitationsdauer nicht ermittelbar. Ersatzweise wird ein vorläufiger Schätzwert gebildet, der sich am empirischen Mittelwert (vgl. Ziff. 2) orientiert. Da sich dieser allerdings aus den genannten Gründen oft nur mittelbar erschließen lässt, wird die proj. Rehabilitationsdauer mittels einer sinnvollen Toleranzspanne angenähert (5.7.3). Die wird für OBS und TBS einerseits und FT andererseits sowie für PrTB und KtrB gesondert durchgeführt.

### 5.7.1 Rehabilitationsdauer

Für die Entwicklung des Modells sind vier Fragen von besonderem Interesse. Wie entsteht die beobachtete Variabilität der individuellen Rehabilitationsdauer? Inwieweit lässt sich anhand der Analyse der Daten zur Rehabilitationsdauer in den vorliegenden Studien (seit 1983) ein Schätzwert für die 'projektierte Rehabilitationsdauer' gesondert nach OBS /TBS vs. FT gewinnen? Ist ein signifikanter Unterschied der Rehabilitationsdauer zwischen der Gewebestruktur OBS /TBS vs. FT erkennbar? Ist ein solcher Unterschied gegebenenfalls auch gesondert nach der Behandlungsform (KtrB bzw. PrTB) erkennbar?

### 5.7.2 Variabilität der 'individuellen Rehabilitationsdauer'

Wie kommt die beobachtete Variabilität der individuellen Rehabilitationsdauer innerhalb einer Studienpopulation zustande?

Die individuelle Rehabilitationsdauer ist bei jedem Patienten natürlicherweise unterschiedlich lang. Innerhalb ein und derselben Studienpopulation streut sie aufgrund kürzerer oder längerer Heilungsverläufe u. a. bedingt durch Nutzungsart, Läsionsgrad oder Alter der Sportpferde (LANGE-CONSIGLIO et al. 2013; PLEVIN und MCLELLAN 2014). Der Heilungsprozess sollte konservativ mit kontrollierter Bewegung unterstützt, aber keinesfalls willkürlich verkürzt werden. Es zeigt sich, dass eine Rehabilitationsdauer mit kontrollierter Bewegung unter 6 M die Inzidenz für ein Rezidiv erhöht (REEF et al. 2001; DYSON 2004; SMITH 2008). Während der Rehabilitation bauen mehrere Heilungsphasen aufeinander auf. Nach der akuten Entzündungsphase (ca. 1-2 Wochen) bildet sich in der Proliferationsphase (ca. 4.-45. Tag) unreifes Gewebe (Kallus). RICHARDSON et al. (2007) schlagen die Proliferationsphase als optimalen Zeitpunkt für die Applikation von prT vor, bevor Fibroblasten vermehrt in die Sehne migrieren und die Vernarbung durch Fibrozytenbildung einsetzt, andererseits aber bereits die Granulation und Vaskularisation des verletzten Sehnenengewebes beginnt; d. h. innerhalb 3-4 Wochen nach der Verletzung. Der Zeitpunkt der Applikation beeinflusst insoweit die Rehabilitationsdauer und daher wahrscheinlich auch den Erfolg von PrTB. Die Kollagenfasern des unreifen Gewebes organisieren sich in der Remodellierungsphase (ca. 45.-120. Tag) zu Bündeln und reifen danach bis zu 6 M mit zunehmender Stärke und Festigkeit (GEBUREK et al. 2015). Experimentell kann der Anstieg von Kollagen I in den ersten 6 M nachgewiesen werden (DAHLGREN et al. 2005); in diesem Zeitraum gewinnt das verletzte Sehnen Gewebe unterstützt durch kontrollierte Bewegung seine ursprünglichen physiologische Eigenschaft sukzessive zurück. In jeder einzelnen Gewebezone der Gliedmaße remodeliert sich die zelluläre Faserstruktur aus unterschiedlichen Kollagen-Typen entsprechend der spezifischen Arbeitsbelastung je nach Nutzungsart, Läsionsgrad und Alter der Sportpferde einzigartig. Das Bewegungstraining sollte umsichtig in einem Maß gesteigert werden, das gewährleistet, dass Entzündungsreaktionen am behandelten Gewebe nicht erneut aufflammen (GILLIS 1997). Ein forciertes Bewegungstraining kann zu Folgeverletzungen und fortschreitender Fibroisierung des geschädigten Gewebes führen. Es erscheint notwendig, den Heilungsfortschritt fortlaufend klinisch zu kontrollieren und u. a. ultrasonographisch zu evaluieren.

### 5.7.3 Vorläufiger Schätzwert für die 'projektierte Rehabilitationsdauer'

#### 5.7.3.1 Analyse der Rehabilitationsdauer zu KtrB bei OBS-TBS vs. FT

Wie erwähnt soll anhand der Analyse der Rehabilitationsdauer der vorliegenden Studien (seit 1983) ein geeigneter Schätzwert für die 'projektierte Rehabilitationsdauer' gewonnen werden. Dabei ist die Frage zu klären, ob ein signifikanter Unterschied der Rehabilitationsdauer zwischen der Gewebestruktur OBS bzw. TBS vs. FT erkennbar ist. Daher sollen zunächst nur Effekte zu KtrB betrachtet werden, so dass Effekte des Behandlungsmittels prT ausgeschlossen sind. Als Datenbasis eignen sich BehG, die Erkrankungen mit KtrB behandeln und möglichst detaillierte Angaben zur Rehabilitationsdauer berichten. Studien vor 2011 geben zur Behandlungsform KtrB bis auf 1A DYSON (2004; N-A = 46) keine Rehabilitationsdauer an; dagegen 5 (der 30) Studien seit 2011: Nr. 1 COUDRY et al. (2014); Nr. 2 TIPTON et al. (2013); Nr. 3 GEBUREK et al. (2015); Nr. 21 LUTTER et al. (2015); Nr. 28 PLEVIN und MCLELLAN (2014). 4 (der 5) Studien (Nr. 1, 2, 21, 28) haben BehG ausschließlich zur Behandlungsform KtrB; 1 (der 5) Studien (Nr. 3) zu KtrB und PrTB. Aus den insgesamt 6 Studien lassen sich somit 6 BehG zu KtrB extrahieren; sie verteilen sich wie folgt auf



die Gewebstruktur: 4 BehG (Nr. 1, Nr. 2, Nr. 3, 1A; in derselben Reihenfolge N-A = 5, 10, 16, 46) bei der OBS sowie 1 BehG (Nr. 21; N-A = 18) bei der TBS und 1 BehG (Nr. 28; N-A = 85) beim FT-Schenkel. Die Fallzahl N-A der 6 BehG hat die Spannweite 1-85 Pferde.

**a) Rehabilitationsdauer zu KtrB bei OBS und TBS.** Die Rehabilitationsdauer der 5 BehG bei OBS bzw. TBS verteilt sich wie folgt: 3-12 *M*; 6 *M* (Dauer des kontrollierten Bewegungsprogramms); 6,75-24 *M*; 9 *M* (Mittelwert); „überwiegend im Bereich“ 12-14 *M* (DYSON 2004). Die 8 Daten werden in aufsteigender Reihenfolge angeordnet: 3; 6; 6,75; 9; 12; 12; 14; 24 *M*. Daraus folgt der empirische Mittelwert 10,8 *M*, die empirische Standardabweichung 6,5 *M*, der Median 10,5 *M* sowie der Interquartilsabstand  $IQR = Q_3 - Q_1 = 13 - 6,4 = 6,6$  *M* und die Spannweite 3-24 *M*. Da es sich bei den Daten teils um eine individuelle Rehabilitationsdauer (Minimum bzw. Maximum der Spannweite), andererseits auch um aggregierte (Mittelwert) wie auch geschätzte Daten („überwiegend im Bereich“) handelt, repräsentieren die zentrale Tendenz (arithm. Mittelwert, Median) und das zugehörige Streuungsmaß (Standardabweichung, Spannweite) nur grobe Orientierungswerte.

**Bewertung.** Das Minimum 3 *M* bzw. Maximum 24 *M* ist eine individuelle Rehabilitationsdauer. Die breite Streuung hängt neben der kleinen Fallzahl (8 Daten) wahrscheinlich mit Unterschieden durch die Einflussfaktoren zusammen wie Läsionsgrad, Altersklasse der Pferde, individuell an die Nutzungsart angepasstes Bewegungsprogramm (VAN SCHIE et al. 2009). Der Interquartilsbereich, der definitionsgemäß 50 % der Daten umfasst, wird hier von der Standardabweichung überdeckt. Sowohl der Mittelwert (10,8 *M*) als auch der Median (10,5 *M*) liegen nahe am Zentrum (9,7 *M*) des Interquartilsbereichs. Legt man eine Normalverteilung der Daten zugrunde, so sprechen alle Indizien dafür den empirischen Mittelwert **10,8 *M*** als einen geeigneten vorläufigen Schätzwert zu KtrB für die '**projektierte Rehabilitationsdauer**' bei Erkrankungen der OBS und TBS anzunehmen.

**b) Rehabilitationsdauer zu KtrB beim FT.** Der FT gliedert sich in drei physiologisch und funktionell-biomechanisch unterschiedliche Gewebestrukturen, proximaler Ursprung, Körper und distale Schenkel. Er ist im Unterschied zu OBS bzw. TBS ein sehniger Muskel mit 2-11 % Muskelfaseranteil, der in den proximalen drei Vierteln des FT-Körpers die meisten und in den Schenkeln noch einzelne Muskelfaserbündel hat (RIJKENHUIZEN et al. 2007; SCHULZE und BUDRAS 2008; BANSE 2011). Die einzige relevante BehG (N-A = 85) beim FT zu KtrB bezieht sich auf Verletzungen leichten bis schweren Grades (LG 4-6) an der Insertion der FT-Schenkel am Gleichbein (Nr. 28, PLEVIN und MCLELLAN 2014) bei Thoroughbred Jährlingen mit der individuellen Rehabilitationsdauer ab 6 *M*. Die Studie deutet an, dass in Einzelfällen ca. 2-4 *M* hinzu kommen.

Greift man Beobachtungen anderer Arbeitsgruppen (RIJKENHUIZEN et al. 2007; SMITH 2016; SCHULZE und BUDRAS 2008) beim FT als Ganzes auf, zeichnet sich beim FT im Vergleich zum Sehnengewebe der OBS bzw. TBS tendenziell eine kürzere Rehabilitationsdauer ab. Beispielsweise wird in der Studie von RIJKENHUIZEN et al. (2007; n = 239) das Erfolgsmerkmal „lahm frei“ einerseits für die Behandlungsform Weidegang ohne Bewegungsprogramm und andererseits für weitere 6 Behandlungsformen kombiniert mit kontrollierter Bewegung bei Erkrankungen von FT-Körper, -Ursprung und -Schenkel untersucht. Die 6 Behandlungsformen sind: Boxenruhe, intra-artikuläre Injektion von Kortikosteroiden, intraläsionale Injektion von Knochenmarkaspirat, Stoßwellentherapie, chirurgische Fasziotomie und Neurektomie der tiefen Äste der lateralen plantaren Nerven nach Bathe. Die individuelle Rehabilitationsdauer bis zur Freiheit von Lahmheit streut für alle Behandlungsformen insgesamt von 3 bis 24 *M*. Es ergibt sich folgende Verteilung: 29,1 % benötigen 3 *M* Rehabilitationsdauer, 35,8 % 4-6 *M*, 17,9 % 7-12 *M*, 7,3 % 13-24 *M*; 9,9 % werden nicht geheilt. Somit benötigen ca. 65 % der Studienpopulation 3-6 *M* zur erfolgreichen Rehabilitation, dagegen nur knapp 25 % länger, nämlich 7-24 *M*. Das Ergebnis von RIJKENHUIZEN et al. (2007) deckt sich mit Beob-

achtungen von SMITH (2016), wonach Erkrankungen der Gewebestrukturen des FT mit 6-9 *M* Rehabilitationsdauer, d. h. im Durchschnitt ca. 7,5 *M*, tendenziell schneller heilen als bei OBS und TBS, die ca. 12 *M* benötigen.

**Bewertung.** Beim FT liegt nur 1 BehG zu KtrB vor. Die empirischen Daten der Studie zur Insertion der FT-Schenkel am Gleichbein (Nr. 28, PLEVIN und MCLELLAN 2014) stützen zwar die Beobachtung, dass **FT**-Erkrankungen tendenziell eine kürzere **Rehabilitationsdauer**, ca. 7,5 *M* haben. Es kann aber nicht ausgeschlossen werden, dass das junge Alter der behandelten Thoroughbred Jährlinge, die kürzere Rehabilitationsdauer begünstigt. Außerdem sind die Daten potenziell mit Bias behaftet, da u. a. der Läsionsgrad und die Altersklasse der Sportpferde unberücksichtigt bleiben. Unsere Entscheidung **ca. 7,5 *M*** als den vorläufigen Schätzwert zu KtrB für die '**projektierte Rehabilitationsdauer**' beim **FT** zu betrachten, ist daher zurückhaltend zu bewerten. Tab 5.7.3.1 fasst das Ergebnis des Vergleichs der Rehabilitationsdauer zu KtrB zwischen OBS-TBS vs. FT zusammen.

Tabelle 5.7.3.1 Vergleich der Rehabilitationsdauer zu KtrB zwischen OBS-TBS vs. FT							
Gewebe- struktur	Anzahl BehG	Anzahl Messwerte	Med	Mw	STD	Spann- weite	vorläufiger Schätzwert der projektierten Rehabilitationsdauer
OBS, TBS	5	8	10,5	10,8	6,5	3 - 24	10,8
FT	1	Erfahrungswert	entfällt	7,5	entfällt	6 – 9	7,5 Erfahrungswert
Rehabilitationsdauer in Monaten; Med Median; Mw Mittelwert; STD Standardabweichung							

#### 5.7.3.2 Analyse der Rehabilitationsdauer zu PrTB bei OBS-TBS vs. FT

Lässt sich der beobachtete Unterschied der vorläufigen Schätzwerte für die 'projektierte Rehabilitationsdauer' zwischen OBS-TBS vs. FT (5.7.3.1) zu KtrB auf eine breitere Datenbasis stellen und bei der Behandlungsform PrTB bestätigen? Dazu ist zu klären, wie der potenzielle Effekt unterschiedlicher prT auf die Rehabilitationsdauer bei OBS-TBS vs. FT einzuschätzen ist. Studien vor 2011 geben zur Behandlungsform PrTB keine Rehabilitationsdauer an. Die 26 (der 30) Studien (2011-2016), die nur mit PrTB behandeln, verteilen sich wie folgt: 11 (der 26) Studien haben BehG nur zur OBS oder TBS (Nr. 5, 6, 9, 10, 12, 14, 17, 18, 19, 20, 22). Weitere 7 (der 26) haben BehG nur zum FT (Nr. 23-27, 29-30). Weitere 7 (der 26) haben BehG zur OBS und zum FT (Nr. 4, 7, 8, 11, 13, 15, 16). Da bei letzteren 7 Studien bis auf Nr. 16 die Rehabilitationsdauer zwischen beiden Gewebestrukturen nicht differenziert, also nur ein und dieselbe Rehabilitationsdauer angegeben wird, fallen diese 6 (von 7) Studien zur Erweiterung der Datenbasis aus. Zusätzlich hat 1 Studie (Nr. 3) zu jeder der beiden Behandlungsformen, KtrB bzw. PrTB, eine BehG zur OBS. (Anmerkung: Aus Gründen der Übersichtlichkeit werden im Folgenden prT von autologen Gewebequellen – anders als allogene und xenogene prT – nicht extra gekennzeichnet).

a) **Potenzieller Effekt von prT bei OBS und TBS.** Wie lässt sich bei OBS und TBS der potenzielle Effekt von prT auf die Rehabilitationsdauer einschätzen? In 13 BehG (aus Studie Nr. 3, 5, 6, 9, 10, 12, 14, 16, 17, 18, 19, 20, 21) werden die Erkrankungen mit 11 unterschiedlichen prT behandelt: Nr. 3 konditioniertes Blutserum; Nr. 5 und Nr. 6 PRP; Nr. 9 Blut-MSK; Nr. 10 Knochenmarkaspirat; Nr. 12 BM-MSK; Nr. 14 Ad-MSK und PRP; Nr. 16 konditioniertes Medium aus allogenen Amniongewebe-MSK; Nr. 17 allogene Ad-MSK und autologes PRP; Nr. 18 allogene Nabelschnurblut-MSK; Nr. 19 und Nr. 20 xenogene ovine Amnion-Epithel-Gewebe-MSK; Nr. 21 allogenes PRP. Da sich die 11 prT auf 13 BehG verteilen, fließen potenzielle therapeutisch unterschiedliche Effekte dieser prT auf die Rehabilitationsdauer annähernd gleich gewichtet in die zentrale Tendenz (empirischer Mittelwert, Median) der Rehabilitationsdauer und die zugehörige Maßzahl für die

Streuung (Standardabweichung, Spannweite) ein.

**Rehabilitationsdauer bei OBS und TBS.** Die 13 BehG ausschließlich zu Erkrankungen der OBS und TBS haben 2-23; 6; 6,75-24; 4-5; 18; 12; 7-9; 12; 3-6; 6; 4; 12; 6-10 Monate Rehabilitationsdauer. Ordnet man die 19 Werte in aufsteigender Reihenfolge 2; 3; 4; 4; 5; 6; 6; 6; 6; 6,75; 7; 9; 10; 12; 12; 12; 18; 23; 24 Monate, so ergibt sich für die Rehabilitationsdauer der Median 6,75 M, die Spannweite 2-24 M, der empirische Mittelwert 9,25 M und die Standardabweichung 6,3 M.

**Bewertung.** Jeder der beiden kürzesten 2 M, 3 M sowie längsten Messwerte 23 M, 24 M verkörpert eine individuelle Rehabilitationsdauer; die breite Streuung erklärt sich neben der kleinen Fallzahl vermutlich durch Unterschiede bei den Einflussfaktoren wie Läsionsgrad, Alter der Pferde, Art des Bewegungsprogramms etc. (VAN SCHIE et al. 2009). Der größte Anteil der Messwerte 73,7 % (14/19) verdichtet sich im Bereich 4-12 M. Die Quartile betragen  $Q_1 = 5$  M und  $Q_3 = 12$  M. Es ist  $IQR = Q_3 - Q_1 = 7$  M. Der Interquartilsbereich umfasst definitionsgemäß 50 % der Messwerte und wird hier von der Standardabweichung 6,3 M (um den Mittelwert 9,25 M) überdeckt. Der Median 6,75 M liegt näher bei  $Q_1$  als bei  $Q_3$ . Der Mittelwert dagegen liegt näher am Zentrum (8,5 M) des Interquartilsbereichs als der Median. Geht man von einer Normalverteilung der Messwerte aus, so deuten alle Indizien auf den empirischen Mittelwert **9,25 M** als dem geeigneten vorläufigen Schätzwert für die '**projektierte Rehabilitationsdauer**' bei mit prT behandelten Erkrankungen der **OBS und TBS**.

**b) Potenzieller Effekt von prT beim FT.** Wie lässt sich beim FT der potenzielle Effekt von prT auf die Rehabilitationsdauer einschätzen? In 9 BehG aus 8 Studien, 7 (der 8) Studien Nr. 23-27 und Nr. 29-30 ausschließlich zum FT, zusätzlich 1 (der 8) Studie Nr. 16 zu FT und OBS (differenziert s. o.), werden die Erkrankungen mit 6 unterschiedlichen prT behandelt. Nr. 16 konditioniertes Medium aus allogenem Amniongewebe-MS; Nr. 25 Ad-MS; Nr. 26 sehnenzell-induzierte MS aus allogenem Blut und allogenes PRP; Nr. 27 natives Knochenmarkaspirat; Nr. 30 mit PRP. In weiteren 3 BehG Nr. 23, Nr. 24, Nr. 29 wird mit PRP behandelt. In Studie Nr. 24 wird in 2 BehG, wie bereits erwähnt, einerseits mit PRP und andererseits mit xenogener porciner UBM behandelt. 4 der 9 BehG behandeln mit PRP, 5 der 9 BehG behandeln mit anderen jeweils verschiedenen prT.

**Rehabilitationsdauer beim FT.** Die 9 BehG ausschließlich zu FT-Läsionen haben 3-6, 2-12, 2-12, 5, 8, ab 6, ab 6, 6, 6 Monate Rehabilitationsdauer. Ordnet man die 12 Werte in aufsteigender Reihenfolge 2, 2, 3, 5, 6, 6, 6, 6, 6, 8, 12, 12 Monate, so ergibt sich der Median 6 M, die Spannweite 2-12 M, der empirische Mittelwert 6,2 M und die Standardabweichung 3,3 M.

Vor einer abschließenden Bewertung dieser Maßzahlen ist zu klären, inwieweit speziell PRP das Ergebnis beeinflusst, da 4 (der 9) BehG mit PRP behandeln. Betrachtet man die Rehabilitationsdauer der 4 BehG 2, 3, 6, 6, 6, 12 Monate, so ergeben sich 6 Werte mit dem Median<sub>4PRP</sub> = 6M, der Spannweite<sub>4PRP</sub> 2-12 M, dem empirischen Mittelwert<sub>4PRP</sub> = 5,8 M und der Standardabweichung<sub>4PRP</sub> = 3,5 M. Betrachtet man die 6 Werte der Rehabilitationsdauer der restlichen 5 (der 9) BehG 2, 5, 6, 6, 8, 12 Monate, so ergeben sich der Median<sub>5prT</sub> = 6 M, die Spannweite<sub>5prT</sub> 2-12 M, der Mittelwert<sub>5prT</sub> = 6,5 M und die Standardabweichung<sub>5prT</sub> = 3,3 M. Der Vergleich der 4PRP-Gruppe mit der 5prT-Gruppe ergibt keinen signifikanten Unterschied, weder bei der zentralen Tendenz (t-Test;  $p = 0,7419$ ) noch bei der Variabilität der Rehabilitationsdauer (F-Test;  $p = 0,9223$ ). So mit verteilen sich potenziell unterschiedliche Effekte der insgesamt 6 prT auf die Rehabilitationsdauer der 9 BehG homogen und fließen etwa gleich gewichtet in die zentrale Tendenz der Rehabilitationsdauer und die zugehörige Maßzahl für die Streuung ein.

**Bewertung.** Die Quartile der 12 Werte für die Rehabilitationsdauer sind  $Q_1 = 4$  M und  $Q_3 = 7$  M. Es ist  $IQR = Q_3 - Q_1 = 3$  M. Der Interquartilsbereich umfasst definitionsgemäß 50 % der Werte und wird hier von der Standardabweichung 3,3 M (um den Mittelwert 6,2 M) überdeckt; außerdem liegt der Median (6 M) ebenso

wie der Mittelwert (6,2 *M*) näher bei  $Q_3$  als bei  $Q_1$ . Geht man von einer Normalverteilung der Daten aus, so sprechen alle genannten Indizien für den empirischen **Mittelwert 6,2 *M*** als dem geeigneten vorläufigen Schätzwert für die '**projektierte Rehabilitationsdauer**' bei mit prT behandelten Erkrankungen des **FT**.

Tab 5.7.3.2 fasst das Ergebnis des Vergleichs der Rehabilitationsdauer zu PrTB zwischen OBS-TBS vs. FT zusammen.

Gewebe- struktur	Anzahl BehG	Anzahl geordnete Messwerte	Med	Mw	STD	Spann- weite	vorläufiger Schätzwert der projektierten Rehabilitationsdauer
OBS, TBS	13	19	6,75	9,25	6,3	2 - 24	9,25
FT	9	12	6	6,2	3,3	2 - 12	6,2
Rehabilitationsdauer in Monaten; Med Median; Mw Mittelwert; STD Standardabweichung							

**Ergebnis.** Wie ist zu PrTB der **Unterschied der beiden vorläufigen Schätzwerte** für die 'projektierte Rehabilitationsdauer' zwischen OBS-TBS vs. FT in Tab 5.7.3.2 zu bewerten? Da zwischen den 13 BehG bei OBS bzw. TBS einerseits und zwischen den 9 BehG beim FT andererseits jeweils die potenziellen Effekte der applizierten prT auf die Rehabilitationsdauer weitgehend homogen und gleichgewichtet verteilt sind, ist zu vermuten, dass ein signifikanter Unterschied der Rehabilitationsdauer zwischen OBS-TBS vs. FT – wenn er tatsächlich existiert – sich in charakteristisch verschiedenen Maßzahlen der zentralen Tendenz niederschlägt. Dabei wird wie bisher idealerweise angenommen, dass sich die Wirkung der Behandlungsmittel, also der 11 unterschiedlichen prT der 13 BehG bei OBS bzw. TBS sowie der 6 unterschiedlichen prT der 9 BehG beim FT, nicht systematisch unterscheidet; und dass sich Einflüsse wie Läsionsgrad oder Altersklasse der Pferde zwischen den 13 BehG bei OBS bzw. TBS und den 9 BehG beim FT zufällig ausgleichen. Zwar unterscheiden sich beide Mittelwerte signifikant ( $p = 0,0432$ ) und deutet die Mittelwertdifferenz 3,05 *M* ( $3,05 = 9,25 - 6,2$ ) auf eine ca. 1,5-fach längere Rehabilitationsdauer bei OBS-TBS vs. FT. Außerdem streut die Rehabilitationsdauer bei OBS-TBS vs. FT signifikant breiter (F-Test;  $p = 0,0146$ ) und ist die Standardabweichung 6,3 *M* vs. 3,3 *M* nahezu doppelt (1,9-mal) so groß; ebenso die Spannweite 2-24 *M* vs. 2-12 *M*. Jedoch ist die Mittelwertdifferenz durch die kleine Fallzahl  $n_1 = 19$  bzw.  $n_2 = 12$  per Signifikanznachweis nicht statistisch gesichert (ca. aktuelle Power 35 %;  $\alpha = 0,05$ ); stattdessen ist die Fallzahl ca.  $n_1 = 56$  bzw.  $n_2 = 36$  erforderlich (aktuelle Power 80,2 %;  $\alpha = 0,05$ ). Somit bleibt zu PrTB die Frage abschließend offen, ob tatsächlich ein Unterschied der Rehabilitationsdauer zwischen der OBS-TBS vs. FT besteht.

#### 5.7.4 'Projektierte Rehabilitationsdauer' gemäß *GFR*-Modell

Wie können die Beobachtungen zu den vorläufigen Schätzwerten (5.7.3) für die Frage nach der 'projektierten Rehabilitationsdauer' zusammengefasst werden? Letztere ergibt summiert zur standardisierten Prüfzeit die standardisierte Studiendauer, die wiederum notwendig ist um Ergebnisse zwischen Studien vergleichen zu können.

**Festlegung der 'Projektierten Rehabilitationsdauer' gemäß *GFR*-Modell.** Die bisherige Analyse aller für die Rehabilitationsdauer relevanten 28 BehG aus 24 (der 30) Studien (2011-2016) und 1 (der 21) Studie (1983-2011) deutet zwar auf signifikante Unterschiede der zentralen Tendenz bei OBS bzw. TBS vs. FT unabhängig von der Behandlungsform KtrB (5.7.3.1) bzw. PrTB (5.7.3.2). Zu KtrB sind dies bei der OBS bzw. TBS vs. FT die Mittelwerte 10,8 *M* vs. 7,5 *M*. Zu PrTB sind dies die Mittelwerte 9,25 *M* vs. 6,2 *M*. Wegen der schmalen Datenbasis kann jedoch der Unterschied der zentralen Tendenz weder zu PrTB noch zu KtrB per Signifikanznachweis statistisch gesichert werden, so dass die gestellte Frage zu beiden Behandlungsfor-

men abschließend offen bleibt.

Allerdings macht die bisherige Untersuchung folgendes deutlich: Da die Rehabilitationsdauer zu PrTB bei der OBS bzw. TBS 2-24 *M* vs. FT 2-12 *M* signifikant breiter streut ( $p = 0,0146$ ), was sich auch für die Rehabilitationsdauer zu KtrB bei der OBS bzw. TBS 3-24 *M* vs. FT 6-9 *M* (Erfahrungswerte) andeutet, besteht die Gefahr, dass die 'projektierte' Rehabilitationsdauer durch die Mittelwerte (bei beiden Behandlungsformen) zu kurz bemessen wird. Die Festlegung darauf würde zu systematischen Fehlern für *GFR* führen. Patienten, die konstitutionell bedingt für die Rehabilitation individuell länger benötigen, würden überproportional zu Fallgruppe A gezählt. Denn bei ihnen reicht die (restliche) Studiendauer insgesamt nicht aus um sie über die volle Länge der standardisierten Prüfzeit auf ein Rezidiv zu kontrollieren. Um einen solchen Bias für das *GFR*-Modell weitgehend auszuschließen wird die zentrale Tendenz um die Toleranzspanne maximal ca. 3 *M* zeitlich sinnvoll verlängert. Diese bemisst sich nach dem Minimum der Standardabweichungen der Rehabilitationsdauer zu KtrB oder PrTB, nämlich 3,3 *M* (PrTB beim FT). Daher wird dieses Minimum (3,3 *M*) adäquaterweise bei beiden Behandlungsformen (PrTB, KtrB) zu den kleinsten Mittelwerten (9,25 *M* bei der OBS-TBS; 6,2 *M* beim FT) addiert. Als 'projektierte Rehabilitationsdauer' gemäß Modell kann somit bei Erkrankungen der OBS bzw. TBS ca. 12 *M* (= 9 *M* + 3 *M*), bei FT-Erkrankungen ca. 9 *M* (= 6 *M* + 3 *M*) definiert werden. – Zwar wäre auch das Maximum der Standardabweichung 6,5 *M* (KtrB bei OBS-TBS) zur Modellbildung der 'projektierten Rehabilitationsdauer' verwendbar, aber dadurch würde die standardisierte Studiendauer (5.7.10) wesentlich verlängert. Inwieweit dies dennoch im Interesse der Forschung sein könnte bleibt zu diskutieren.

Tab 5.7.4 fasst die bisherigen Ergebnisse der Analyse aller für die Rehabilitationsdauer relevanten BehG aus 24 (der 30) Studien (2011-2016) und 1 (der 21) Studie (1983-2011) zusammen.

Tabelle 5.7.4 <b>Projektierte Rehabilitationsdauer gemäß <i>GFR</i>-Modell</b>								Projektierte Reha- bilitationsdauer gemäß <i>GFR</i> - Modell
Gewebe- struktur	Anzahl BehG	Anzahl geordnete Messwerte	Med	Mw	STD	Spann- weite	Schätz wert vorläufig	
KtrB								
OBS, TBS	5	8	10,5	10,8	6,5	3-24	10-11	12
FT	1	Erfahrungswert	entfällt	7,5	entfällt	6–9	6-9	9
PrTB								
OBS, TBS	13	19	6,75	9,25	6,3	2 - 24	9,25	12
FT	9	12	6	6,2	3,3	2 - 12	6,2	9
Rehabilitationsdauer in Monaten; Med Median; Mw Mittelwert; STD Standardabweichung								

### Exkurs: Bewegungsprogramme nach GILLIS (1997) und SMITH (2016)

GILLIS (1997) entwickelt gestützt auf 2800 Fälle von Läsionen bei OBS bzw. TBS und FT, die entweder mit Weidegang allein oder mit Boxenruhe kombiniert mit kontrollierter Bewegung behandelt werden, ein 12 monatiges Rehabilitationsprogramm. Die Notwendigkeit eines solchen Programms wird dadurch unterstrichen, dass nur die mit Weidegang allein behandelten Pferde signifikant weniger zur Arbeit zurückkehren. Das Bewegungsaufbautraining ('Führen; im Schritt Reiten; Traben; Galoppieren; bis zum Wettkampftempo gesteigerter Galopp') kann individuell an den Grad der Verletzung ('mild bzw. leicht, moderat bzw. mäßig, schwer') sowie den ultrasonographisch evaluierten Heilungsfortschritt ('schwach, angemessen, gut') des Pferdes angepasst werden. Nach GILLIS (1997) sollten Übermüdung und Überbelastung unbedingt vermieden werden, da das verletzte Gewebe am Anfang des Heilungsprozesses noch relativ schwach ist und erst über Monate hin

wieder stärker wird. Der Grad der Verletzung erweist sich für das Erfolgsmerkmal 'Rückkehr zur Arbeit' (66,7 %; 59,3 %; 44,4 %) als signifikanter Einflussfaktor. Leichte Verletzungen folgen einem verkürzten Rehabilitationsplan von 3-5 *M*. Durchschnittlich benötigen erfolgreiche Fälle 8-9 *M* bis zur Rückkehr zur vollen Arbeitsbelastung. Die elementaren Funktionen des Rehabilitationsprogramms zur Förderung der Heilung bestehen einmal darin, die Bewegungs- und Gleitfähigkeit (Elastizität) des muskuloskeletären Apparates aufrecht zu erhalten, zum anderen das Aufflammen von Entzündungen zu verhindern. Folgende Studien der vorliegenden Metaanalyse schließen sich dem Bewegungsprogramm von GILLIS (1997) an: TIPTON et al. 2013; O'MEARA et al. 2010; DYSON 2004; CREVIER-DENOIX et al. 2010; DYSON (2004) empfiehlt eine Wiederaufnahme des Galoppierens nicht vor 12-16 *M* oder erst, nachdem die Sehnenheilung durch eine US-Untersuchung diagnostisch sicher belegt ist. Auch RANTANEN et al. (2011) legen in ihrem Entwurf für ein 7-stufiges Bewegungsaufbauprogramm den Schwerpunkt auf die individuelle Anpassung je nach Nutzungsart und beschreiben die Vorgehensweise detailliert.

Auf der Grundlage übereinstimmender empirischer Daten und orthopädischer Erkenntnisse zu Heilungsprozessen bei Sehnenläsionen schlägt SMITH (2016) vor, wie ein geeignetes Bewegungsprogramm aufgebaut werden kann. Er betont die Notwendigkeit der flexiblen proaktiven Anpassung der kontrollierten Bewegung: Verletzungen einiger Gewebestrukturen verkürzen das Bewegungsprogramm, andere erfordern spezifische Bewegungsübungen oder im Gegenteil gar keine; z. B. strikte Ruhe bis zur Heilung bei Fesselgelenkserkrankungen. Neben Trainingsuntergrund (hart, weich, schlüpfrig etc.) sowie orthopädischer Hufbeschlag führt er weitere 5 wesentliche Faktoren an und erläutert dies an Beispielen:

- 1) **Gewebestruktur.** Sehnenverletzungen in tieferen Gewebebereichen ('mid-substance') benötigen ca. 12 *M* lang ein Bewegungsprogramm. FT-Läsionen heilen dagegen schneller in ca. 6-9 *M*. Verletzungen des USB der TBS benötigen ein nur 3-6 *M* langes Bewegungsprogramm; BEERTS et al. (2013; Tab 4.2.1 Nr. 22) bestätigen die Rehabilitationsdauer 4 *M*.
- 2) **Schwere der Läsion.** Bei Verletzungen, die mit dem Verlust der Unterstützung des Fesselgelenks einhergehen, ist strikte Ruhe bis zur Heilung zu verordnen.
- 3) **Heilungsrate.** Das Bewegungsprogramm kann je nach klinischem Fortschritt beschleunigt oder verlangsamt werden.
- 4) **Typ der Sehnenverletzung.** Eine Verletzung am Sehnenrand innerhalb der Sehnenscheide stellt hohe Anforderungen an die optimale Wahl des geeigneten Bewegungsregimes; da nur der äußere Anteil der Sehnenquerschnittsfläche betroffen ist, ist die Belastungssteigerung der Sehne nicht das vorrangige Ziel. Bei tiefer im Gewebe liegenden Sehnenläsionen erscheint eine frühe progressive Mobilisation (s. u.) besser geeignet als exzessive Bewegung, da letzteres entzündliche Reaktionen provozieren kann. Bei einblutenden Sehnenläsionen und chirurgischen Wundexzisionen wiederum sollte die Bewegung in frühen Phasen der Heilung minimal gehalten werden. Laut HEINTEL (2013) ist der verringerte Sehnenquerschnitt an der verletzten Stelle, die sog. 'Sehnentaille', eine hohe Gefahrenzone für eine Sehnenruptur. Da in einem solchen Fall die Kraft dämpfende und Kraft leitende Funktion der Sehne eingeschränkt ist, sollte die Wiederaufnahme und Steigerung der Belastung umsichtig erfolgen.
- 5) **Disziplin** bzw. Nutzungsrichtung. Nach SMITH (2013) tolerieren bspw. Traber bestehende Läsionen des Sehnenapparates und FT besser und können trotz solcher Verletzungen das Training oder Rennen fortsetzen, während die meisten Pferde anderer Disziplinen in einem solchen Fall noch dringend die Fortführung des Rehabilitationsprogramms benötigen. Aus unserer Sicht sollten allerdings Tierschutz relevante Aspekte nicht vernachlässigt werden.

Die konstitutiven Elemente für ein optimiertes Bewegungsprogramm nach SMITH (2016) werden hier kurz skizziert. Verschiedene Gangarten (Schritt, Trab, Galopp) belasten den muskuloskeletären Apparat biomechanisch unterschiedlich hoch: Im Vergleich zum ruhenden Körpergewicht ist die Belastung des Bewegungsapparats beim Schritt um 60 %, beim Trab um 90-100 %, beim Galopp um 130-140 % höher. Daher beginnt

das Bewegungsprogramm zunächst in der Entzündungsphase mit **Stallruhe und kontrollierter passiver Mobilisation**. Die Entzündungsphase geht in die reparative Phase der Sehnenheilung über, die sich durch Angiogenese und Invasion von Fibroblasten auszeichnet und gewöhnlich 1-2 Wochen dauert. In dieser Phase sollte frühzeitig mit der **progressiven Mobilisation** begonnen werden, was oft zu einer Verbesserung der Lahmheit führt. In der folgenden Remodellierungsphase 3-4 *M* (und länger) sollten die Bewegungsübungen auf dem untersten Anforderungs-Level beginnen und graduell bis täglich maximal 30-45 Minuten 'nur Schritt' länger werden, um zu vermeiden dass das gerade entstehende, noch unreife und schwache Ersatzgewebe nicht erneut verletzt wird, sondern genügend Zeit hat, strukturell fester und belastbarer zu werden. Da das verletzte Sehnengewebe in keinem Fall wieder vollständig hergestellt wird (reparatio, keine restitutio ad integrum), bleibt eine Differenz zu den ursprünglichen physiologischen Gewebeeigenschaften, was ein Rezidiv begünstigt. Ziel ist es, diese Differenz zu minimieren. Deshalb sollte das kontrollierte Behandlungsgregime fortlaufend (erstmal 4-7 Tage nach der Verletzung; danach mindestens alle 3 *M*) per US-Untersuchung evaluiert werden, um bereits erste Anzeichen (semiquantitativer Vorher-Nacher-Vergleich: Echogenität und parallele Faserstruktur; quantitativer CSA, MIZ und T-Lesion %) eines Rezidives erkennen zu können und gegebenenfalls ein zu rasch gesteigertes Trainingstempo oder zu hohes Trainingslevel rechtzeitig zu drosseln.

- Exkurs-Ende

#### 5.7.5 Prüfzeit

Die Prüfzeit ist notwendig zur nachhaltigen Kontrolle des Erfolgsmerkmals EM1 'Nach erfolgreicher Rehabilitation kehren die Sportpferde zur Arbeit auf mindestens dem ursprünglichen Leistungsniveau zurück' (5.5.1) auf ein Rezidiv. Bisher folgt man Erfahrungswerten bei der Prüfzeit. Um jedoch *GFR* zwischen Studien vergleichbar zu machen, ist eine gleich lange Prüfzeit erforderlich, da unabhängig von der Behandlungsform (PrTB, KtrB) mit länger werdender Prüfzeit die Zahl der Rezidivfälle steigt (LUTZ 2011). Deren Länge bestimmt maßgeblich, wie nachhaltig der Behandlungserfolg zu bewerten ist und somit, ob *GFR* valide ist, d. h. allgemeine Schlussfolgerungen für die umfassende Grundgesamtheit gezogen werden können. Daher werden die Studien daraufhin analysiert, inwieweit die Prüfzeit für das Modell standardisiert werden kann.

#### 5.7.6 Kriterien der Standardisierung der Prüfzeit

Anders als bei Suche nach dem geeigneten Näherungswert für die 'projektierte Rehabilitationsdauer' entfällt bei der Analyse der Prüfzeit der Studien zwar die Unterscheidung der beiden Behandlungsformen PrTB und KtrB, jedoch nicht zwischen der Gewebestruktur OBS-TBS einerseits und FT andererseits. Die Aufgabe besteht darin, für zwei konkurrierende Anforderungskriterien einen praxisnahen Kompromiss für ein tragfähiges Konzept des Modells zu finden. Daher wird die vergleichende Analyse der Studien unter zwei Fragestellungen durchgeführt.

- 1) Wie lange sollte die Prüfzeit mindestens bemessen sein, damit differenziert nach OBS-TBS vs. FT anhand des Erfolgsmerkmals EM2 'Nach erfolgreicher Rehabilitation bleibt der Patient in der Prüfzeit auf dem erworbenen Leistungsniveau rezidivfrei' die Rezidivrate *R* valide bestimmt werden kann?
- 2) Wie lange sollte die Prüfzeit maximal bemessen sein, damit die Studiendauer im Interesse der Forschenden und Patientenbesitzer bzw. Trainer für die Durchführung klinischer Studien geeignet erscheint?

#### 5.7.7 Analyse der Prüfzeit bei OBS-TBS und FT

Nimmt man zu den Prüfzeiten der vorliegenden Studien in 2011-2016 die der relevanten Studien 1983-2011

aus der Metaanalyse von GEBUREK und STADLER (2011b) hinzu, so ergibt sich folgender Überblick:

**a) Erkrankungen des FT.** Aus 8 Studien vor 2011 sind 8 BehG extrahierbar, die nur FT-Erkrankungen behandeln (DYSON 2000; PERSONETT 1983; RIJKENHUIZEN et al. 2007; WASELAU et al. 2008; CASTELLIJNS et al. 2011; KÜNNECKE et al. 2008; HERTHEL 2001; ROSENBROCK et al. 2004). Für 8 Studien ab 2011 (Nr. 23-30) trifft dies ebenfalls zu. Allerdings scheiden 4 (der 8) Studien wieder aus; denn 3 (der 4) Studien vergleichen Rennleistungen (HALL et al. 2013; GARETT et al. 2013; PLEVIN und MCLELLAN 2014); die letzteren beiden bei Thoroughbred Jährlingen, mit den Prüfzeiten 24-60, 10-18, 6-48 Monate. Diese Prüfzeiten sind maßgeblich dadurch bestimmt, dass die Jährlinge an der dreijährigen bzw. vierjährigen Rennsaison teilnehmen (vgl. Exkurs „Rennleistung und Prüfzeit“). In der vierten Studie (RICH et al. 2014) erscheint die ungewöhnlich lange Prüfzeit von 7,7 Jahren (92 *M* s. u.) als Sonderfall bzw. Ausreißer. Somit verbleiben noch vier relevante (der 8) Studien ab 2011 (Nr. 23, 24, 26, 30); ihre Prüfzeiten sind: 6-24, 17, 12, 12 *M*. Die 8 Studien vor 2011 haben die Prüfzeiten: 3, 3, 3, 6, 6-18, 12, 12-36, 36 *M*. Zusammen ergibt das geordnet 15 Prüfzeiten: 3, 3, 3, 6, 6, 6, 12, 12, 12, 12, 17, 18, 24, 36, 36 *M*.

**Bewertung:** Inwieweit können die Kriterien der Standardisierung erfüllt werden? Es erscheint plausibel, dass die Prüfzeit mindestens so lange sein sollte wie die Rehabilitationsdauer. Orientiert man sich an der 'projektierten Rehabilitationsdauer' ca. 9 *M* bei FT-Erkrankungen (5.7.4) als Referenzwert, so erscheint der kleinste Wert 3 *M* als Prüfzeit zu kurz für eine valide Prüfung auf ein Rezidiv. Auch der größte Wert 36 *M* erscheint als Prüfzeit nicht geeignet; denn summiert zur 'projektierten Rehabilitationsdauer' resultiert daraus eine vergleichsweise überdurchschnittlich lange Studiendauer. Wie sind die 15 Werte statistisch deskriptiv zu beurteilen? Eliminiert man die beiden Werte 3 *M*, 36 *M* (5 Daten), so haben die verbleibenden 10 empirischen Prüfzeiten die Spannweite 6-24 *M*, den Median 12 *M*, den Mittelwert 12,5 *M* und die Standardabweichung 5,9 *M*. Der Inter-Quartils-Abstand  $IQR = Q_3 - Q_1$  ist ein gegenüber Ausreißern (d. h. besonders große oder kleine Werte) robuster Schätzer und umfasst definitionsgemäß 50 % aller Werte; es ist  $Q_1 = 6$  *M* und  $Q_3 = 17$  *M* und  $IQR = 17 - 6 = 11$  *M*. Da der Median 12 *M* und der Mittelwert 12,5 *M* nahe dem Zentrum 11,5 *M* des IQR liegt, ist die zugrunde liegende Verteilungsfunktion der Prüfzeit wahrscheinlich symmetrisch, was auf eine Normalverteilung deutet.

**b) Erkrankungen der OBS und TBS.** Vor 2011 untersuchen 8 (der 21) Studien nur die OBS, keine die TBS. Nach 2011 untersuchen 15 (der 30) Studien (Nr. 1, 2, 3, 5, 6, 9, 10, 12, 14, 17, 18, 19, 20, 21, 22) Erkrankungen nur der OBS oder TBS. 3 (der 8) Studien und 5 (der 15) Studien machen keine Angabe zur Prüfzeit. Es verbleiben 5 relevante Studien vor 2011 (DYSON 2004; SMITH 2008a; SMITH 2008b; GODWIN et al. 2011; PACINI 2007) und 10 relevante Studien nach 2011 (Nr. 1, 2, 3, 6, 9, 10, 17, 19, 20, 21) mit insgesamt (in derselben Reihenfolge) 15 Angaben zur Prüfzeit: 12, 12, 24, 24, 24, 24, 6-36, 24-48, 24, 36, 24, 24, 6, 6, 4-72 *M*. Geordnet ergibt das 18 Daten 4, 6, 6, 6, 12, 12, 24, 24, 24, 24, 24, 24, 24, 24, 36, 36, 48, 72 *M*.

**Bewertung:** Inwieweit können die Kriterien der Standardisierung erfüllt werden? Wählt man die 'projektierte Rehabilitationsdauer' ca. 12 *M* bei Erkrankungen der OBS oder TBS (5.7) als Referenzwert, so erscheinen die Werte 4 *M* und 6 *M* als Prüfzeit zu kurz für die valide Prüfung auf ein Rezidiv. In der Studie Nr. 21 LUTTER et al. (2015) streut die Prüfzeit mit der Spannweite 4-72 *M*. Der Wert 4 *M* erklärt sich wahrscheinlich dadurch, dass die retrospektive Studie aus dem verfügbaren Datenbestand (2000-2010) selektiv nur Patienten untersucht, die einem Bewegungsprogramm unterzogen werden, das auf 6 *M* ausgelegt ist; jedoch sollte die Prüfzeit sinnvollerweise mindestens so lang sein wie die Rehabilitationsdauer. Der kurze Prüfzeitraum 6 *M* erscheint nur in Studie Nr. 2 (TIPTON et al. 2013) gerechtfertigt, da er sich speziell auf Cutting Horses bezieht, deren periphere Verletzungen der OBS von leichtem bis mittleren Läsionsgrad LG 4-5 rechtzeitig behandelt erfahrungsgemäß innerhalb von 4 *M* heilen. Die Werte 48 *M* und 72 *M* erscheinen als Prüfzeit nicht geeignet, da infolge die Studiendauer zu sehr ausgedehnt wird. Wie sind die statistisch deskriptiv erhobenen



Kennzahlen der 18 Werte zu beurteilen? Eliminiert man die Werte 4, 6, 48, 72 *M*, so haben die verbleibenden 12 empirischen Prüfzeiten den Median 24 *M*, Mittelwert 24 *M*, die Standardabweichung 7,2 *M* und die Spannweite 12-36 *M*. Der Inter-Quartils-Abstand  $IQR = Q_3 - Q_1$  beträgt  $IQR = 24 - 24 = 0$  *M*. Da der Median sowohl mit  $Q_3$  als auch  $Q_1$  und dem Zentrum (24 *M*) des IQR zusammenfällt, sind 50 % der Werte mit dem Median identisch; dies betrifft auch den Mittelwert. Daher ist die zugrunde liegende Verteilungsfunktion der Prüfzeit wahrscheinlich annähernd symmetrisch, was auf eine Normalverteilung deutet.

#### 5.7.8 Vergleich der Prüfzeit zwischen OBS-TBS vs. FT

Die statistisch deskriptive Untersuchung der Studien in 5.7.7 führt zu folgenden Kennzahlen der absoluten Verteilung des Merkmals Prüfzeit (in Monaten) für OBS-TBS vs. FT (Tab 5.7.8):

Tabelle 5.7.8 Absolute Verteilung der Prüfzeit bei OBS-TBS vs. FT										
FT	—		Q <sub>1</sub> = Min	Median	Mw = 12,5	Q <sub>3</sub> = 17	Max	—		
Prüfzeit	3, 3, 3	-	6, 6, 6	12, 12, 12, 12	--	18	24	36, 36	-	-
Prüfzeit	-	4	6, 6, 6	12, 12	24, 24, 24, 24, 24, 24, 24, 24, 24			36, 36	48	72
OBS /TBS	—			Min	Mw = Median = Q <sub>1</sub> = Q <sub>3</sub> = 24			Max	—	
Prüfzeit in Monaten; Mw Mittelwert; Q <sub>i</sub> Quartil; Kennzahlen bzgl. der bereinigten Verteilung										

Die Werte 3 *M* bzw. 36 *M* in der FT-Gruppe und 4, 6 bzw. 48, 72 *M* in der Sehnengruppe sind als Prüfzeit entschieden zu kurz bzw. zu lang und werden gestrichen. Welche Schlussfolgerungen für die Standardisierung der Prüfzeit lassen sich mithilfe der statistisch induktiven Untersuchung aus dem Vergleich der beiden bereinigten Verteilungen zwischen OBS-TBS vs. FT ziehen? Die empirischen Mittelwerte 12,5 bzw. 24 *M* unterscheiden sich signifikant ( $p = 0,0003$ ). Sie repräsentieren die durchschnittliche Prüfzeit, die in klinischen Studien seit 1983 als Erfahrungswerte jeweils verwendet werden. Die empirischen Standardabweichungen 5,9 bzw. 7,2 *M* sind fast gleich groß. Die empirische Varianz der Prüfzeit unterscheidet sich somit erwartungsgemäß nicht signifikant ( $F = 1,5178$ ;  $p = 0,2702$ ). Zieht man auch den Inter-Quartils-Abstand  $IQR = Q_3 - Q_1$  als quantitatives Maß für die Streuung heran, so wird ein interessanter Unterschied deutlich. Bei OBS-TBS-Erkrankungen ist  $IQR = 0$  *M*; daher fallen 50 % der Prüfzeiten mit dem Median zusammen. Die Prüfzeit bei OBS-TBS-Erkrankungen ist konsistent. Der Befund deutet darauf, dass der Heilungsprozess bis zur Rückkehr zur Arbeit erwartungsgemäß bei OBS und TBS viel Zeit beansprucht und die Patienten auch lange (24 *M*) beobachtet werden müssen, um sicher klinisch diagnostizieren zu können, ob die Erkrankung 'völlig' geheilt ist. Beim FT dagegen lässt die Prüfzeit 12,5 *M* zusammen mit der vergleichsweise großen Variabilität  $IQR = 11$  *M* vermuten, dass die Rehabilitationsdauer – neben Nutzungsart, Läsionsgrad und Alter der Patienten – von dem erkrankten biologischen Gewebeabschnitt (proximaler Ursprung, Körper, distale Schenkel) abhängt und je nachdem kürzer oder länger ausfallen kann. Danach bemisst sich die geeignete Prüfzeitdauer. Ein Blick auf die Studien könnte dies bestätigen: 7 (der 12) Studien zum FT untersuchen den FT-Ursprung; 5 (der 7) haben erwartungsgemäß die kürzere Prüfzeit 3, 3, 3, 12, 12 *M*; nur 2 (der 7) haben dagegen 6-24 *M*, 17 *M*. Weitere 2 (der 12) Studien untersuchen den FT-Körper bzw. die FT-Schenkel; sie haben – zumindest was die FT-Schenkel betrifft – die lange Prüfzeit 36 *M*, 12-36 *M*. Die übrigen 3 (der 12) Studien untersuchen den FT ohne Berücksichtigung der Lokalisation; die zugehörige Prüfzeit 6, 12, 6-18 *M* streut breit um den Durchschnittswert ca. 10,5 *M*. Die Vermutung lässt sich jedoch aufgrund der schmalen Datenbasis der 15 Prüfzeiten weder quantitativ signifikant belegen noch anhand aller Studien direkt vergleichen, da nur diese 4 Studien in 2011-2016 die Rehabilitationsdauer berichten.

### 5.7.9 Prüfzeit gemäß *GFR*-Modell

Als wichtiges Teilergebnis aus der bisherigen Analyse der Prüfzeit für das *GFR*-Modell ist zu erkennen, dass sich die Entscheidung, wie lange beim FT die standardisierte Prüfzeit zu bemessen ist, nicht allein nach der zentralen Tendenz (Mittelwert 12,5 *M*) richten darf, sondern auch die Variabilität der Prüfzeit zu berücksichtigen hat. Die objektiven Bemessungskriterien für die Prüfzeit orientieren sich gesondert nach OBS-TBS vs. FT einerseits am Mittelwert (24 *M* bei OBS-TBS vs. 12,5 *M* beim FT) und andererseits am Quartil  $Q_3$  (24 *M* bei OBS-TBS vs. 17 *M* beim FT).  $Q_3$  ist ein repräsentativer Vergleichspunkt der Variabilität, da er 75 % der Prüfzeiten aller BehG umfasst. Wählt man daher  $Q_3 = 17$  *M* als Prüfzeit, so wird berücksichtigt, dass die Rehabilitationsdauer am proximalen Ursprung tendenziell kürzer und an den distalen Schenkeln vergleichsweise länger sein kann. Der rechnerische Wert 17 *M* wird zur einfacheren Handhabung auf 18 *M*, d. h. 1,5 Jahre, sinnvoll gerundet.

Die Analyse der 25 relevanten (der 51) Studien seit 1983 ergibt somit als standardisierte Prüfzeit zur Kontrolle der GF von Sportpferden nach Rehabilitation bei Erkrankungen der OBS oder TBS gemäß *GFR*-Evaluationsmodell 24 *M* und bei FT-Erkrankungen 18 *M*. Tabelle 5.7.9 fasst das Ergebnis zur Prüfzeit gemäß *GFR*-Modell differenziert nach OBS bzw. TBS vs. FT zusammen.

Tabelle 5.7.9 <b>Standardisierte Prüfzeit gemäß <i>GFR</i>-Modell bei OBS-TBS vs. FT</b>									
	Anzahl Studien / BehG	Anzahl Prüfzeiten	Median	Mw	STD	$Q_1$	$Q_3$	Spannweite	Standardisierte Prüfzeit gemäß <i>GFR</i> -Modell
OBS, TBS	13	12	24	24	7,2	24	24	12-36	<b>24</b>
FT	12	10	12	12,5	5,9	6	17	6-24	<b>18</b>
Zeitangabe in Monaten; $Q_i$ Quartil; STD Standardabweichung; BehG Behandlungsgruppe; Mw Mittelwert									

### 5.7.10 Standardisierung der Studiendauer gemäß *GFR*-Modell

Aufgrund der Untersuchungen zur Rehabilitationsdauer und Prüfzeit in 5.7.1-9 kann die Studiendauer im größtmöglichen Konsens mit der bisherigen Forschungspraxis standardisiert werden. Bei FT-Erkrankungen beträgt die standardisierte Studiendauer **2,25 Jahre** = (9 *M* projektierte Rehabilitationsdauer + 18 *M* Prüfzeit); bei Erkrankungen der OBS und TBS beträgt sie **3 Jahre** = (12 *M* projektierte Rehabilitationsdauer + 24 *M* Prüfzeit). Einerseits erscheint diese Länge der Studiendauer im Interesse von Patientenbesitzern (Trainern) und Forschenden für die Durchführung einer klinischen Studie geeignet; andererseits kann die Rezidivrate *R* bei genügend großer Fallzahl innerhalb der definierten Studiendauer valide bestimmt werden. Auf Basis der damit ermittelten empirischen Gebrauchsfähigkeitsrate *GFR* können allgemeine Schlussfolgerungen für die umfassende Grundgesamtheit gezogen und weiterführende Hypothesen zur Entwicklung des Modells generiert werden. Tabelle 5.7.10 fasst das Ergebnis zur Standardisierung der Studiendauer gemäß *GFR*-Modell zusammen.

Tabelle 5.7.10 <b>Standardisierte Studiendauer gemäß <i>GFR</i>-Modell</b>			
Gewebestruktur	Projektierte Rehabilitationsdauer	Standardisierte Prüfzeit	Standardisierte Studiendauer
OBS, TBS	12	24	<b>36 = 3 Jahre</b>
FT	9	18	<b>27 = 2,25 Jahre</b>
Zeitangabe in Monaten			

### 5.7.11 Zusammenhang zwischen Zunahme der Prüfzeit und Abnahme der *GFR*

Mit zunehmender Prüfzeit ist zu erwarten, dass bei gleichen Bedingungen die Anzahl *C* der Rezidivfälle naturgemäß steigt, wogegen *GFR* kleiner wird. Um den Zusammenhang statistisch belegen zu können, werden BehG benötigt, die sich nur in der Prüfzeit unterscheiden, aber in den übrigen Einflussfaktoren auf *GFR* übereinstimmen. In die engere Wahl kommen die 11 Paare (5.4.2.2). 3 (der 7) Paare, nämlich 3P, 5P, 6P sind in Gewebestruktur, prT, Altersklasse, Läsionsgrad und Nutzungsart sowie paar-intern in der Prüfzeit 24 *M* vs. 12 *M* weitgehend identisch. Allerdings ist die Fallzahl der 6 BehG (in gleicher Reihenfolge N-A = 5+5; 2+2; 5+5) sehr klein, sodass damit kein aussagekräftiges Ergebnis erzielt werden kann. Außerdem ist nur bei Paar 6P in der BehG mit der längeren Prüfzeit *GFR* tatsächlich kleiner. Bei 3P, 5P dagegen ist *GFR* bei beiden BehG paar-intern gleich groß.

Da bei Paar 11P in beiden BehG (4A bzw. 5A; LEPPÄNEN et al. 2009a bzw. 2009b) ausschließlich Traber behandelt werden, insbesondere also die Nutzungsart übereinstimmt, sei bezüglich des vermuteten Zusammenhangs ein Hinweis erlaubt.  $GFR_{4A} = 51 \% [0,37-0,65]$  der BehG (N-A = 53) mit der längeren Prüfzeit 24 *M* ist erwartungsgemäß signifikant ( $p = 0,0181$ ) kleiner als  $GFR_{5A} = 74 \% [0,55-0,88]$  der BehG (N-A = 31) mit der kürzeren 12 *M*. Insoweit belegt 11P den vermuteten Zusammenhang. Da jedoch die erforderliche Fallzahl fehlt um die Differenz statistisch zu sichern und 1P keine Angaben zum Alter und Läsionsgrad macht, erscheint jeder weitergehende Vergleich nicht angemessen.

#### Exkurs: Rennleistung und Prüfzeit

Eine besondere Gruppe bilden klinische Studien, die Rennleistungen bei Pferden vergleichen. Sie beziehen Überlegungen zur Länge des geeigneten Prüfzeitraums systematisch in ihre Fragestellung ein (GIBSON et al. 1997; O'MEARA et al. 2010; RENZI et al. 2013; HALL et al. 2013; GARETT et al. 2013; PLEVIN und MCLELLAN (2014); HU et al. 2014; RUSSEL et al. 2016).

Stellvertretend sei hier O'MEARA et al. (2010) genannt; die Autoren beobachten, dass die festgelegte Anzahl von Rennen, die erreicht werden muss, allein kein geeignetes Maß ist, um Erfolge unterschiedlicher Behandlungsformen bei Rennpferden objektiv vergleichen zu können. Vielmehr ist dafür die biomechanische Leistung der Pferde geeignet, wie sie durch eine definierte Anzahl an Trainingseinheiten oder Wettkämpfen während einer definierten Zeitspanne erbracht wird. In einer kontrollierten klinischen Studie ( $n = 802$ , Kontrollgruppe aus Unverletzten; Fallgruppe aus Verletzten; 3 Behandlungsformen von Verletzungen der OBS: Insulin like growth factor-type 1, bar firing, desmotomy of superior check ligament) kommen die Autoren (per Matching bzgl. Alter der Pferde, Geschlecht, Trainingsstall) bei Rennpferden (Flach-, Hürden-, Jagd-Rennpferde) zu dem Schluss, dass nach 1 Jahr Rehabilitation **mindestens fünf aufeinander folgende Rennen** im Zeitraum von zwei Jahren notwendig sind um signifikante Unterschiede zwischen der Rennleistung der Gruppen nachweisen zu können. Allerdings berichten O'MEARA et al. (2010), dass die nach 1 Jahr Rehabilitation und danach 2 Jahre Wettkampf- oder Trainingszeitraum ermittelte Rezidivrate *R* noch besser geeignet ist um unterschiedliche Behandlungserfolge bei Verletzungen der OBS valide zu beurteilen. Dies deckt sich mit den Überlegungen des *GFR*-Modells. RUSSELL et al. (2016) greifen das Ergebnis von O'MEARA et al. (2010) auf und empfehlen als 'Goldstandard' für die Beurteilung von Rennleistungen das Erfolgsmerkmal: 'Ein Rennpferd kehrt zur ursprünglichen Rennleistung für mindestens 2 Jahre oder zu 5 Starts rezidivfrei zurück'. Vergleicht man die Erfolgsmerkmale 'Rückkehr zu Rennen' (RENZI et al. 2013; COUDRY et al. 2014; GUERCIO et al. 2015) bzw. 'Teilnahme an mehr als einem Rennen' (ZUFFOVA et al. 2013; HALL et al. 2013) sowie 'Starts bis Ende der 3- bzw. 4-jährigen Rennsaison' speziell bei Thoroughbred Jährlingen (PLEVIN und MCLELLAN 2014; GARETT et al. 2013) mit dem Erfolgsmerkmal bei RUSSEL et al. (2016), so zeigt sich: (1) Letzteres Erfolgsmerkmal erscheint insbesondere dann, wenn die geforderten mindestens 5

Rennen in weniger als 2 Jahren (Trainer-Ehrgeiz) absolviert werden, nicht mit dem Merkmal 'rezidivfrei für mindestens 2 Jahre bei ursprünglicher Rennleistung' vergleichbar. (2) Außerdem ist die Rate  $E$  für das Erfolgsmerkmal 'Teilnahme an mehr als einem Rennen' mit der Rate  $E$  des Erfolgsmerkmals 'Absolvieren von mindestens fünf Rennen' nicht vergleichbar, denn bereits zwei absolvierte Rennen zählen als Erfolg.

Die klinische Studie RUSSELL et al. (2016) untersucht per US diagnostizierte akute (weniger als 4 Wochen alte) 'core lesions' - mit nahezu keinem Echo bis anechogen - an der OBS bei 34 Trabern und 71 Flachrennpferden, die zur Erstbehandlung vorgestellt werden. Das Behandlungsmittel ist einerseits Desmotomie des Unterstützungsbandes der OBS kombiniert mit Knochenmarkaspirat (89 Fälle) oder andererseits nur Knochenmarkaspirat (16 Fälle). Nach der Rehabilitation (1,5 a) erreichen 61,8 % (55/89) aus der Desmotomie- und 50 % (8/16) aus der Knochenmarkaspirat-Gruppe das Erfolgsmerkmal. - Exkurs-Ende

## 5.8 Ausblick

### 5.8.1 Das gemeinsame Effektgrößen-Paar der Form ( $GFR$ / Rehabilitationsdauer)

Wie bereits eingangs (2.2) erwähnt ist zu erwarten, dass die Behandlung mit einem prT die Rehabilitationsdauer verkürzt, wenn der Behandlungserfolg einem anderen prT überlegen ist. Aus ethischen Erwägungen des Tierschutzes wird vorausgesetzt, dass sich bei einer Verkürzung der Rehabilitationsdauer die Gewebequalität nicht verschlechtert und infolge sich  $GFR$  im Vergleich zu KtrB verringert. Dies führt zu der Frage, inwieweit die Rehabilitationsdauer als quantitative Effektgröße verwendet werden kann um idealerweise bei gleichen Bedingungen Behandlungserfolge zwischen verschiedenen prT valide zu vergleichen. Das Modell eröffnet die Möglichkeit die Rehabilitationsdauer mit dem empirischen Wert  $1,5GFR$  bzw.  $2GFR$  zu kombinieren. Es entsteht das 'gemeinsame Effektgrößen-Paar' der Form ( $GFR$  / Rehabilitationsdauer der Studienpopulation). Damit können Unterschiede von PrTB gegenüber KtrB sowie zwischen verschiedenen prT durch eine kürzere oder längere Rehabilitationsdauer quantitativ verglichen werden. Zwei 'gemeinsame Effektgrößen-Paare' sind mittels Rehabilitationsdauer vergleichbar, wenn der Unterschied zwischen den beiden Effektgrößen  $GFR_1$  und  $GFR_2$  durch die erforderliche Fallzahl der Studienpopulation per Signifikanznachweis statistisch gesichert ist (z. B. Power 80 %;  $\alpha = 0,05$ ). Bei den folgenden Fallbeispielen ist die Bedingung durch  $N-A = 131$  in jeder BehG für die aktuelle Power 80,1 % und aktuelles  $p = 0,0185$  erfüllt. Damit können die 3 grundlegenden Fallkonstellationen veranschaulicht werden.

#### Drei Fallkonstellationen sind möglich:

**1) Fall:** Bei beiden 'gemeinsamen Effektgrößen-Paaren' ist entweder **(1a)** die Rehabilitationsdauer gleich lang und  $GFR$  verschieden groß oder **(1b)**  $GFR$  gleich groß und die Rehabilitationsdauer verschieden lang.  
**Beispiel 1a:** Erstes Paar (85 % / 10 M); zweites Paar (70 % / 10 M). Bei gleich langer Rehabilitationsdauer 10 M ist die Chance auf GF bei der ersten Behandlungsform wahrscheinlich 2,4-fach ( $OR = 2,4$ ) höher. **Beispiel 1b:** Erstes Paar (70 % / 12 M); zweites Paar (70 % / 8 M). Bei gleicher Erfolgchance ( $OR = 1$ ) ist die Rehabilitationsdauer bei der ersten Behandlungsform gegenüber der zweiten tendenziell 1,5-mal (12 / 8) länger.

**2) Fall:** Zwischen beiden Paaren ist  $GFR$  und Rehabilitationsdauer verschieden groß und die paar-interne Erfolgsausrichtung gleichsinnig. **Beispiel 2:** Erstes Paar (85 % / 6 M); zweites Paar (70 % / 10 M). Bei dieser Fallkonstellation ist leicht zu entscheiden. Bei der ersten Behandlungsform ist sowohl  $GFR$  größer als auch die Rehabilitationsdauer kürzer und bei der zweiten sowohl  $GFR$  kleiner als auch die Rehabilitationsdauer länger. Die Chance auf GF ist bei Paar 1 wahrscheinlich 2,4-fach höher und die Rehabilitationsdauer 0,6-mal kürzer.

**3) Fall:** Zwischen beiden Paaren ist *GFR* und Rehabilitationsdauer verschieden groß und die paar-interne Erfolgsausrichtung gegensinnig. **Beispiel 3:** Erstes Paar (85 % /10 *M*); zweites Paar (70 % /6 *M*). Zwar erbringt die 0,6-fach kürzere Rehabilitationsdauer der zweiten Behandlungsform den absoluten Zeitvorteil 4 *M*. Der Zeitvorteil führt aber zu Ungunsten der Patienten zu der wahrscheinlich *OR* = 0,4-fach geringeren Chance der GF. Im konkreten Beispiel ist die erste Behandlungsform vorzuziehen, da aus ethischen Erwägungen der Schutz des Tierwohls vorgeht.

### 5.8.2 Die Norm-*GFR* der Behandlungsform KtrB

Bereits GEBUREK und STADLER (2011b) weisen darauf hin, dass die Rezidivrate *R* zu KtrB eine zentrale Bedeutung sowohl im Vergleich mit Rezidivraten vorhandener wie auch neu entwickelter prT erlangen könnte. *GFR* zu KtrB könnte die Funktion einer Norm-*GFR* übernehmen, um damit andere *GFR* relativ zu einem absoluten 'natürlichen Nullpunkt' valide quantitativ zu vergleichen. Ein Ziel veterinärmedizinischer Forschung könnte es daher sein, einen „KtrB-Goldstandard“ – differenziert nach Einflussfaktoren (Gewebestruktur, Nutzungsart, Läsionsgrad, Altersklasse) – zu entwickeln. Inwieweit können dazu die vorliegenden BehG zu KtrB beitragen? Bisher wird KtrB (*n* = 5) zwar ein einziges Mal in einer RCT (Nr. 3; vgl. Tab 5.1a) zur positiven Kontrolle von PrTB (*n* = 10) eingesetzt, aber mit kleiner Fallzahl und somit wenig präzisiertem Ergebnis. Gemäß *GFR*-Modell können im Zuge prospektiv geplanter Metaanalysen BehG zu KtrB auch von nicht randomisierten – kontrollierten oder unkontrollierten – klinischen Studien erzeugt und ihre Daten gepoolt werden. So können zwar nicht Ergebnisse auf höchster Evidenz-Stufe, jedoch mit steigender Fallzahl immer präzisere Schätzwerte für *GFR* gewonnen werden. Die vorliegende Metaanalyse (2011-2016) findet für eine solche Norm-*GFR* zu KtrB bei der OBS für LG 4 den Näherungswert  $2GFR = 57,5\%$ ; ihm liegen 3 poolbare BehG mit der Fallzahl *N-A* = 60 Sportpferde zugrunde (5.4.4).

### 5.8.3 Modellevaluierung und Weiterentwicklung

Durch Ergebnisse zukünftiger Studien können die Maßzahlen für die 'projektierte Rehabilitationsdauer' oder die 'Prüfzeit' möglicherweise besser angenähert werden. Diskutiert werden kann, welche Vor- und Nachteile es hat, die Prüfzeiten für OBS, TBS (24 *M*) und FT (18 *M*) einheitlich auf 24 *M* zu standardisieren. Zukünftige Studien werden klären, ob die definierte Prüfzeit ausreicht um Effektdifferenzen  $d_{GFR}$  zwischen verschiedenen PrTB hinreichend genau zu evaluieren. Diagnostische Parameter für CAS und FAS können mit fortschreitender Entwicklung der US-Analyse-Technik mit höherer Sensitivität und Spezifität gemessen werden. Zusätzlicher Forschungsbedarf besteht um mit dem Modell den Einfluss einzelner Faktoren zu beurteilen, etwa labortechnische Herstellungsverfahren von MSC oder Blutprodukten (PRP: Plasmapherese-Filter, Sedimentation, Zentrifugation einmal /mehrmals).

Bei der Berechnung der Rezidivrate einer Studienpopulation wird für jeden Patienten nur betrachtet, ob während der Prüfzeit ein Rezidiv beobachtet wird oder nicht. Es wird nicht berücksichtigt, wann dieses Rezidiv aufgetreten ist. Diese Betrachtung ist im Allgemeinen sehr vereinfacht, da die Zeitspanne bis zum Ereignis außer Acht gelassen wird. Werden die Rezidivraten einer Studienpopulation bei (idealerweise) gleichen Bedingungen zu aufeinanderfolgenden Zeitpunkten oder in Zeitperioden (z. B. monatlich) erfasst, so bilden die Beobachtungen Zeitreihen. Die Analyse solcher Zeitreihen könnte anders als bei Querschnittsanalysen Trends in Form glatter Funktionen der Zeit erkennen lassen z. B. abhängig von prT, Alter der Pferde, Gewebestruktur und Prognosen erlauben.

Im Folgenden wird zusammenfassend diskutiert, ob und inwieweit die eingangs aufgestellten Arbeitshypothesen bestätigt werden können.

## 6.1 Bisherige Bewertungspraxis

Die Hypothese eingangs lautet: Die metaanalytische 'gepoolte Reanalyse' kann mithilfe eines evidenzbasierten einfachen Evaluationsmodells mit der neuen quantitativen Effektgröße 'Gebrauchsfähigkeitsrate' *GFR* die Ergebnisse der bisherigen Studien (1983-2016) zu PrTB und KtrB unabhängig von den Ergebnissen der Primärstudien neu bewerten. Diese Hypothese findet sich zwar weitgehend bestätigt. Allerdings ist es bei rund der Hälfte der 100 extrahierten BehG nicht möglich *E*, *R* bzw. die Werte für die Variablen A, B, C differenziert nach der Gewebestruktur aus den Studien zu bestimmen. Da die Kenntnis der Schwächen der bisherigen Bewertungspraxis zum besseren Verständnis des Modells beiträgt, werden 7 Beispielfälle diskutiert und Denkanstöße für ihre Revision gegeben. Im Folgenden bedeutet N die Gesamtanzahl der zu Beginn einer Studie behandelten Patienten.

### 6.1.1 Rezidivrate

Manche Studien bestimmen die Rezidivrate *R*, indem sie die Zahl der Rezidivfälle *C* auf die gesamte Studienpopulation *N* oder auf eine Teilmenge beziehen. Der Begriff der Rezidivrate *R* erscheint dadurch unscharf und der Behandlungserfolg ist nicht vergleichbar.

**Beispielfall 1.** Die Studien von LEPPÄNEN et al.(2009b), GODWIN et al. (2012), ROSENBROCK et al. (2004), RICCO et al. (2013), RENZI et al. (2013), LUTZ (2011) beziehen die Anzahl *C* der Rezidivfälle in der Form  $R^* = C / N$  auf *N*.  $R^*$  ist zwischen Studien nur vergleichbar, wenn nach Behandlung mit PrTB kein Fall A oder B auftritt. Dagegen erscheint *R* eindeutig definiert, wenn *C* exakt auf die Anzahl der regelkonform behandelten Patienten mit erfolgreicher Rehabilitation bezogen wird, d. h.  $R = C / (N - A - B)$ . Ist A oder B größer Null, folgt  $R^* < R$ ;  $R^*$  lässt den Behandlungserfolg günstiger erscheinen als er tatsächlich ist.

Die Studie von LEPPÄNEN et al. (2009b) berechnet für *N* = 53 behandelte Fälle (meist Trabrennpferde; Gewebestruktur: 'OBS, TBS, USB und FT gemischt')  $R^* = 13,2 \%$  ( $7 / 53$ ); d. h. es sind absolut gezählt *C* = 7 Rezidivfälle. *A*<sub>1</sub> = 7 Verletzungsfälle, die zur Behandlung als Rezidivfälle vorgestellt werden, erleiden nach Behandlung wieder an derselben Gliedmaße eine Verletzung, was der Wirkung des Behandlungsmittels nicht allein ursächlich zugeschrieben werden kann. Sie werden von LEPPÄNEN et al. (2009b) – korrekterweise – rechnerisch eliminiert, um das Ergebnis des Behandlungserfolgs nicht zu verzerren. Weitere *A*<sub>2</sub> = 8 (von 53) Fälle werden laut Studie rechnerisch eliminiert, da sie sich nach der Behandlung an anderen als den behandelten Gliedmaßen neu verletzen. Insgesamt berichtet die Studie *A* = 14 Fälle; vermutlich wird ein Fall in der Studie doppelt gezählt. Patienten, die das Erfolgsmerkmal nicht erreichen und der Fallgruppe B zugeordnet werden müssen, treten in der Studie nicht auf, d. h. es ist *B* = 0. Es verbleiben *N*-*A*-*B* = 39 (= 54-14-0) regelkonform behandelte Fälle mit erfolgreicher Rehabilitation, von denen *C* = 7 Fälle ein Rezidiv erleiden. *R* gemäß Modell beträgt somit  $R = C / (N - A - B) = 7 / (53 - 14 - 0) = 7 / 39 = 17,9 \%$  und ist ersichtlich bedeutend größer als  $R^* = 13,2 \%$ .

**Beispielfall 2.** In der einfach (bzgl. Bewerter) verblindeten RCT (*n* = 20) von GEBUREK et al. (2016) besteht die Versuchs- und Kontrollgruppe (OBS: autologes PRP vs. Kochsalzlösung) jeweils aus 10 Pferden. Die Studie betrachtet den Anteil an gebrauchsfähigen Pferden, der noch 24 *M* nach Behandlung kombiniert mit einem kontrollierten Bewegungsprogramm (6 *M*) auf dem ursprünglichen Leistungsniveau arbeitet, als

Maß für den Behandlungserfolg; er wird mit 60 % (6 /10) für die PRP-Gruppe bzw. mit 50 % (5 /10) für die Kontrollgruppe angegeben. Die Werte fallen jedoch gemäß *GFR* anders aus, da Fälle A und C auftreten: In jeder Gruppe tritt 1 Rezidivfall auf, somit ist jeweils  $C = 1$ . Ebenfalls in jeder Gruppe stirbt 1 Pferd und wird 1 Pferd zurückgezogen aus Gründen, die mit dem Behandlungsmittel kausal nicht zusammenhängen; somit ist jeweils  $A = 2$ . In jeder Gruppe lassen sich 2 Pferde nach der Rehabilitationsphase nur auf einem geringeren Leistungsniveau als ursprünglich reiten; somit ist jeweils  $B = 2$ . Allerdings entwickelt in der PRP-Gruppe zusätzlich ein Pferd eine Sehnenverletzung an der kontralateralen Gliedmaße; nach dem *GFR*-Modell erhöht sich dort die Anzahl von  $A = 2$  auf  $A = 3$  um potenzielle Störfaktoren, etwa aufgrund eines Zusammenhangs mit der ursprünglichen Verletzung, auszuschließen. Damit berechnet sich in der PRP-Gruppe  $GFR_{PRP} = (N - A - B - C) / (N - A) = (10 - 3 - 2 - 1) / (10 - 3) = 4 / 7 = 57,1 \%$  und in der Kontrollgruppe  $GFR_{Kontroll} = (10 - 2 - 2 - 1) / (10 - 2) = 5 / 8 = 62,5 \%$ . Die laut Studie angegebenen Behandlungserfolge 60 % vs. 50 % verhalten sich dazu gegensätzlich. Bei anderen Variablenwerten A, B, C könnte der Unterschied noch widersprüchlicher ausfallen.

### 6.1.2 Rehabilitationsdauer, Prüfzeit und Erfolgsmerkmal

Für die valide Beurteilung des Behandlungserfolgs müssen Erfolgsmerkmal, projektierte Rehabilitationsdauer der Studienpopulation und Prüfzeitraum zueinander passen. Dies gelingt vielen Studien nicht zufriedenstellend.

#### a) Das Erfolgsmerkmal 'recovery rate' mit der festgelegten 'Rehabilitationsdauer ein Jahr'

**Beispielfall 3.** Um das regenerative Potenzial von kultivierten MSC aus autologem Fettgewebe zu messen, wählen LEPPÄNEN et al. (2009a) als Erfolgsmerkmal (EM) die 'recovery rate' (Wiederherstellungsrate) nach einem Jahr; das ist der Anteil der 'recovery population', dem es nach Behandlung gelingt, bereits innerhalb eines Jahres wieder an Rennen teilzunehmen. Während einer Rehabilitationsdauer von 8 *M* kehren rund 50 % der 'recovery population' zu Rennen zurück, während eines Jahres rund 85 %. Die Erfolgsrate (bzgl. EM) nach einem Jahr ist somit  $E' = 85 \%$ . Jedoch fehlt die Kontrolle der Erfolgsrate  $E'$  mittels Rezidivrate  $R$ . - Auf ähnliche Weise untersuchen COUDRY et al. (2014) das Erfolgsmerkmal 'Rückkehr zu Rennen' bei 16 an der OBS verletzten Trabern, die in der Studie als Kontrollgruppe eingesetzt und mit KtrB behandelt werden. Nach durchschnittlich 8 *M* kehren  $E' = 50 \%$  (8/16) der Pferde zurück, nach durchschnittlich 9 *M* sind es  $E'' = 87,5 \%$  (14/16). Die Prüfzeit und somit auch die Kontrolle der Erfolgsraten  $E'$ ,  $E''$  mittels Rezidivrate  $R$  fehlen auch in dieser Studie.

#### b) Das Erfolgsmerkmal (EM): 'Rückkehr zum ursprünglichen Leistungsniveau nach einem beschleunigten Rehabilitationsprogramm von 5 Monaten und anschließend rezidivfrei für mindestens 1 Jahr'

**Beispielfall 4.** Die retrospektive klinische Studie von RICH et al. (2014) untersucht anhand der Datenbank (2005-2010) einer Pferdeklinik die Patientendaten von 83 am FT (52 vorn, 28 hinten, 3 vorn und hinten) mit Ad-MSC behandelten Sportpferden (vor allem Western und English Performance; Alter der Pferde:  $Mw = 9$  a;  $STD = 4$  a; Spannweite 1-19 a). Es werden ausschließlich Einriss-Verletzungen behandelt. Schwere der Verletzung und CSA verteilen sich wie folgt: 9 Pferde *mild* 5-25 % CSA; 57 Pferde *mäßig* 26-50 % CSA; 26 Pferde *schwer* 52-75 % CSA; 1 Pferd *last resore* 76-99 % CSA. Die Studie untersucht 176 FT-Läsionen; das ist – nach Ansicht der Autoren – die höchste Anzahl in bisherigen Studien zum FT. Vor Behandlung haben 27 (der 83) Patienten bereits andere Verletzungen. Die Läsionen (64 Pferde akut; 19 Pferde chronisch: mindestens 3 *M* vor Behandlung) verteilen sich beim FT wie folgt: proximaler FT-Ursprung (72 Läsionen bei 30 Pferden), FT-Körper (40 Läsionen bei 6 Pferden), FT-Schenkel (64 Läsionen bei 15 Pferden) sowie „mehrere

Segmente“ (32 Pferde). Die Ad-MSC-Behandlung ist kombiniert mit einem Bewegungsaufbauprogramm von maximal 5 M Dauer (2,5-3 M gemäßigt Training; sobald in dieser Zeit geführter Trab möglich ist, folgt 2 M ein gesteigertes Bewegungsaufbautraining im Sattel). Wie einige andere Studien beziehen auch RICH et al. (2014) die Erfolgsrate  $E'$  (bzgl. EM') undifferenziert auf alle 3 Gewebestrukturen des FT. Die klinische Datenbank, die der Studie zu Grunde liegt, ermöglicht es die Leistungsentwicklung aller Pferde 13-92 M nach Rückkehr zum ursprünglichen Leistungsniveau zu beobachten. Es wird die beachtlich hohe Erfolgsrate  $E' = 84,3 \%$  bezüglich der gesamten Studienpopulation 83 Patienten berichtet. Das Erfolgsmerkmal unterscheidet sich vom GFR-Modell. Die willkürliche Festlegung der individuellen Rehabilitationsdauer auf 5 M deutet auf die Gefahr eines Selektionsbias. Da der Behandlungserfolg bevorzugt für solche Pferde ermittelt wird, bei denen der Heilungsprozess rascher voranschreitet als beim Rest. Allgemeine Schlussfolgerungen für die umfassende Grundgesamtheit sind so nicht bzw. nur eingeschränkt möglich. In der Studie treffen allerdings 3 Faktoren zusammen, die den raschen Heilungsfortschritt innerhalb der kurzen Rehabilitationsdauer 5 M begünstigen: **(1)** 80 % aller Pferde haben Läsionen von einem *milden* (11 % Pferde) bzw. *mäßigen* (69 % Pferde) CSA-Grad. **(2)** 77 % aller Pferde haben akute Läsionen; nach COWLES (2000), den RICH et al. (2014) zitieren, benötigen akute Verletzungen des FT 1-4 M Rehabilitationsdauer. **(3)** DYSON (2000) stellt fest, dass kontrollierte Bewegung bei Läsionen am proximalen FT-Ursprung der Vordergliedmaße erheblich erfolgreicher ist als an der Hintergliedmaße. Dies deckt sich mit der Beobachtung, dass bei RICH et al. (2014) 44,5 % der Studienpopulation solche Läsionen haben; denn 62,6 % (52 /83) der Pferde haben Läsionen an der Vordergliedmaße und 71,1 % (37 /52) davon am proximalen FT-Ursprung. Dagegen haben nur 20,4 % der Studienpopulation Läsionen am proximalen FT der Hintergliedmaße; denn 33,7 % (28 /83) der Studienpopulation haben Läsionen an der Hintergliedmaße und davon 60,7 % (17 /28) am proximalen FT. Auch in der Studie (n = 20; OBS) von GEBUREK et al. (2016) begünstigt der hohe Anteil von peripheren Läsionen geringen Grades und von diffusen Läsionen die Heilungsprognose mit einem beschleunigten kontrollierten Bewegungsprogramm von nur 6 M Rehabilitationszeit: 11 (der 20) Pferde (55 %) haben periphere Läsionen geringen Grades, 6 (der 20) Pferde diffuse und nur 3 (der 20) Pferde haben Läsionen schweren Grades (core lesions). TIPTON et al. (2013) zeigen, dass die überwiegende Zahl von Cutting-Horses, bei denen periphere Sehnenverletzungen der OBS typisch sind, rechtzeitig behandelt bereits nach einer mittleren Rehabilitationsdauer von 4 M zum vollen Training zurückkehren.

### c) Unscharfe Definition des Erfolgsmerkmals

Ungünstig für die Bewertung des Behandlungserfolgs wirkt sich eine unscharfe Definition des Erfolgsmerkmals aus. Bereits RICH et al. (2014) bemängeln, dass die Vergleichbarkeit von Behandlungsergebnissen oft an der unterschiedlichen Definition des Erfolgsmerkmals in den Studien scheitert: z. B. 'wieder lahm frei basierend auf einer einzigen Nachuntersuchung'; 'Rückkehr zum vollen Arbeitseinsatz ohne Nachkontrolle'; 'Rückkehr zu einem von den Besitzern angestrebten bestimmten Leistungsniveau'.

**Beispielfall 5.** In einer klinische Studie (Kortikoid; Hyaluronsäure) mit insgesamt 118 Fällen werden N = 18 Fälle mit Erkrankungen der TBS mit KtrB behandelt (LUTTER et al. 2015). Als Erfolgsmerkmal wird die 'Rückkehr zur Arbeit auf drei Leistungsniveaus nach Rehabilitation (6 M) am Ende der Prüfzeit (4-72 M)' definiert. *Excellent* bedeutet die Rückkehr auf das ursprüngliche Leistungsniveau. *Good* bedeutet die Rückkehr auf ein niedrigeres Leistungsniveau, was im GFR-Modell der Fallgruppe B entspricht. *Poor* bedeutet keine Rückkehr zur Arbeit oder das Ausscheiden des Patienten aus der Studienpopulation, was vermutlich Fallgruppe A, aber auch A+C entsprechen könnte, da sich die Studie zu Rezidivfällen nicht äußert. Alternativ kann C = 0 vermutet werden. Für das Merkmal *poor* werden 11 Fälle angegeben. Weiterhin unterscheidet die Studie bei der Auswertung der Daten nicht zwischen *excellent* und *good*, sondern gibt dafür zusammen 5 Fälle an. Aufgrund dessen kann GFR nicht bestimmt werden.



### 6.1.3 Lückenhafte Präsentation der Quelldaten

Oft werden in Studien die Quelldaten nicht so umfänglich präsentiert, wie es für die interessierende Fragestellung der Metaanalyse nötig wäre. Teils fehlen Angaben zu A, B, C bzw. R, E. Oder die Fragestellung der untersuchten Studie zielt in eine andere Richtung.

**Beispielfall 6.** Die klinische Studie von SCALA et al. (2014) hat  $n = 150$  Pferde; unterschiedliche Rassen; für 99 Pferde beträgt die protokollierte Prüfzeit mindestens 12 M; Alter der Pferde  $Mw = 9,3$  a, Spannweite 2-23 a. Behandelt werden die Läsionen mit PRP kombiniert mit kontrollierter Bewegung (Bewegungstraining 3 M) an folgenden Gewebestrukturen: OBS (45), TBS (8), Gelenkbänder (6), FT (35) und USB der TBS (5). Die Behandlung wird per US (3., 6., 9. Woche) bis zur 'völligen klinischen Heilung' evaluiert. Die Fragestellung der Studie zielt darauf, ob sich der Behandlungserfolg zwischen 4 verschiedenen Applikationsprotokollen unterscheidet: (1) am 1. Tag intraläsional kombiniert mit periläsional; (2) am 1. Tag intraläsional und am 15. und 30. Tag periläsional; (3) am 1. und am 30. Tag intraläsional; (4) am 1. Tag intraläsional gefolgt von einer Kombinationsbehandlung aus Stoßwelle und Wärme (Tecar Therapie) bis zum 25. Tag. Der Therapie-Erfolg wird nach 3 Stufen bewertet: (1) *improvement* (12 %, 12 Pferde) bedeutet, das Pferd erreicht die 'völlige klinische Heilung', nimmt aber seine sportliche Aktivität auf einem niedrigeren Leistungsniveau auf als ursprünglich. (2) *success* (81 %, 80 Pferde) bedeutet, das Pferd erreicht die 'völlige klinische Heilung' und beginnt innerhalb von maximal 6 M Rehabilitationsdauer wieder auf dem ursprünglichen Leistungsniveau mit sportlicher Aktivität. (3) *failure* (7 %, 7 Pferde) bedeutet, die 'völlige klinische Heilung' wird nicht erreicht. Die Studie präsentiert folgende Ergebnisse ohne nähere Angaben: Die Altersklassen (jung/alt) der Pferde unterscheiden sich hinsichtlich der erreichten Therapie-Erfolgs-Stufen signifikant (ANOVA;  $p = 0,05$ ). Die 4 Behandlungsformen unterscheiden sich bezüglich der erreichten Stufen nicht signifikant. Auch nicht zwischen unterschiedlichen Gewebestrukturen (ANOVA;  $p = 0,05$ ); was laut Studie wahrscheinlich mit der gleich hohen therapeutischen Wirkung von PRP bei allen Gewebestrukturen erklärt werden kann, da selbst einige Fälle mit kompletter Sehnenruptur geheilt werden.

Die Studienstatistik weist zwar die Gesamtzahl der Pferde aus, die eine bestimmte Stufe erreichen (*success* 80 Pferde etc.), jedoch nicht wie sich diese Anzahl auf die verschiedenen Gewebestrukturen verteilt. Alle Versuche B und C nach der Gewebestruktur differenziert abzuschätzen enden unbefriedigend; z. B. ergibt sich bei der OBS (45 Pferde)  $0 \leq B \leq 19$  bzw.  $0 \leq C \leq 8$ . Zur Verifizierung der Verteilung der Häufigkeit der Altersstufe und des Läsionsgrad liegen die Quelldaten nicht vor. Es wird nur knapp berichtet, dass alle behandelten Pferde einen Lahmheitsgrad 'mindestens 2' haben sowie, dass Ort und Schweregrad der Läsionen mittels eines longitudinalen und transversalen US-Scans (3., 6., 9. Woche nach Behandlung) diagnostiziert werden.

Im Prüfzeitraum (mind. 12 M) erleiden 8 Pferde (10 %, 8 /80) der Stufe *success* ein Rezidiv. 3 (der 8) Pferde, die die Stufe *success* erreicht haben und im Prüfzeitraum (mind. 12 M) ein Rezidiv erleiden, erreichen nach einer zweiten PRP-Applikation erneut die Stufe *success*. Damit wird die interessante Frage aufgeworfen, wie sich die Behandlung eines Rezidivs innerhalb der Prüfzeit mit dem gleichen prT auf den Behandlungserfolg auswirkt; die Frage kann hier nicht weiter verfolgt werden.

**Beispielfall 7.** Bei manchen Studien kann die Rezidivrate R für einzelne behandelte Gewebestrukturen nur geschätzt werden. Die Studie von LANGE-CONSIGLIO et al. (2013a) behandelt 95 Sportpferde: 44 (der 95) mit BM-MSC an FT (10), TBS (2) und OBS (32) sowie 51 (der 95) mit allogenen Amniongewebe-MSC an FT (14), TBS (3) und OBS (34). In der Knochenmark-Gruppe ist A = 0. Innerhalb der Rehabilitationsdauer 4-12 M nach Behandlung kehren 39 (der 44) Pferde zum ursprünglichen Leistungsniveau zurück; d. h. für alle Gewebestrukturen betrachtet ist  $B = 44 - 39 = 5$ . Da nur diese Angabe vorliegt, kann für die Knochenmark-Gruppe bei der OBS nur vermutet werden, dass maximal  $B_{OBS} = 5$  ist. Für die OBS liegt  $E_{OBS} = (N - A - B) / (N - A)$  somit zwischen 84 % (27/32) und 100 % (32/32). Während der zweijährigen Prüfzeit erleiden 23

% (10/44), also absolut gezählt 10 Pferde ein Rezidiv, wovon 7 Fälle sich auf die OBS beziehen, wie die Auswertung der Tabelle der Primärdaten ergibt. d. h. die Rezidivrate  $R_{OBS}$  für die Gewebestruktur OBS liegt zwischen maximal  $R_{OBS-max} = C / (N-A-B) = 7 / (32-0-0) = 26 \%$  und mindestens  $R_{OBS-min} = 7 / (32-0-5) = 22 \%$ , somit ist  $GFR_{OBS-min} = 62,5 \%$  und  $GFR_{OBS-max} = 78 \%$ .

## 6.2 Diversifizierung der prT

Die Hypothese, aus den Studien inhaltlich homogene BehG mit kleiner Fallzahl zu größeren BehG poolen und Unterschiede des Behandlungserfolgs  $GFR$  mittels Power-Analyse durch die erforderliche Fallzahl per Signifikanznachweis statistisch sichern zu können, lässt sich nur an wenigen BehG bestätigen. Zum einen führt die bisherige Bewertungspraxis (6.1) dazu, dass  $GFR$  nur für 50 von insgesamt 100 aus den Studien 1983-2016 extrahierbaren BehG bestimmt werden kann. 11 (der 50) BehG kombinieren die Gewebestruktur und sind nicht vergleichbar. Zum anderen trägt die steigende Diversifizierungsrate der prT in den Studien im Verlauf von 2001-2011 und 2011-2016 maßgeblich dazu bei, dass von den restlichen 39 (der 50) BehG nur für 6 BehG mit der Fallzahl  $30 \leq N-A \leq 145$  aussagekräftige Vergleiche möglich sind. Denn die Diversifizierungsrate, d. h. die Zahl unterschiedlicher prT bezogen auf die Gesamtzahl der Studien zu PrTB steigt von 38 % (6/16) auf 73,1 % (19/26) um ca. das Doppelte (1,9-fach); das entspringt dem wachsenden Interesse an allogenen prT, die ab 2011 erstmals appliziert werden. Von den 19 unterschiedlichen prT (bezüglich der 100 BehG) verbleiben 10 prT (bezüglich der 50 BehG), die sich bei den 39 (der 50) BehG auf 5 Gewebestrukturen (OBS, TBS, FT-Ursprung, -Körper, -Schenkel) verteilen.

4 (der 39) BehG gehören zu KtrB; aus 3 (der 4) BehG zu KtrB kann 1 BehG bei der OBS mit der Fallzahl  $N-A = 60$  gepoolt werden.

Zu PrTB ist für weitere 6 (der 39) BehG das Poolen möglich, so dass aus je 2 BehG (sog. Paare) 3 größere BehG mit der Fallzahl  $42 \leq N-A \leq 145$  entstehen.

Zu diesen insgesamt 4 gepoolten BehG kommen 2 einzelne BehG mit  $30 \leq N-A$  hinzu.

Zwischen diesen 6 BehG sind 15 (6 über 2) paarweise Vergleiche (pV) möglich; allerdings sind nur 7 (der 15) pV adäquat; sie werden in Tab. 5.4.4.3 präsentiert; die restlichen 8 (der 15) pV unterscheiden sich in Gewebestruktur und prT und sind daher nicht vergleichbar.

## 6.3 Hohe bzw. niedrige Evidenz der Ergebnisse

Die Hypothese, dass mithilfe des  $GFR$ -Modells replizierbare Ergebnisse von hoher interner Evidenz erzielt werden können, wird – bedingt durch die 6.1 und 6.2 dargelegten Gründe – nur für 3 (der 7) pV (6.2) bestätigt: Für Amniongewebe-MS (LANGE-CONSIGLIO et al. 2013a) bei der OBS kann die Power-Analyse im Vergleich mit KtrB, Knochenmark-MS und PRP den Unterschied der GF mittels Effektdifferenz  $d_{GFR}$  durch die erforderliche Fallzahl per Signifikanznachweis statistisch sichern. Gleichzeitig belegt der Nachweis die Erwartung der equinen Regenerationsmedizin, dass PrTB wahrscheinlich KtrB überlegen ist.

### 6.3.1 Amniongewebe-MS

Die größer Chance  $OR$  ( $7,2 \leq OR \leq 11,4$ ) für GF bei Behandlung mit allogenen Amniongewebe-MS bei der OBS erscheint gegenüber Knochenmark-MS (GODWIN et al. 2012; LANGE-CONSIGLIO et al. 2013a), PRP (ABELLANET et al. 2009; GEBUREK et al. 2016) und KtrB (DYSON 2004; TIPTON et al. 2013; GEBUREK et al. 2015) statistisch gesichert (aktuelle Power > 90 %; aktuelles  $p < 0,025$ ). Das Ergebnis ist vielversprechend. Die größere Plastizitäts- und Proliferationskapazität von Amniongewebe-MS, die bereits LANGE-CONSIGLIO et al. (2013a) gegenüber Knochenmark (Paar 1;  $2GFR = 68,8 \%$ ; Tab 5.4.4.3) beob-

achten, wird durch die größere GF mittels  $2GFR = 94,1 \%$  (2-Jahres-Gebrauchsfähigkeitsrate) gegenüber PRP ( $2GFR = 60,4 \%$ ) und KtrB ( $2GFR = 57,5 \%$ ) quantitativ belegt. Amniongewebe vereint im Vergleich zu anderen Gewebequellen mehrere Vorteile. Das prT kann tiefgefroren gelagert werden. Es steht in 'Realttime' innerhalb 6-15 Tage zur Injektion in das verletzte Sehnen- oder Bandgewebe zur Verfügung, also in der Proliferationsphase bevor im US irgendwelche klinisch relevanten Veränderungen auftreten (LANGE-CONSIGLIO et al. 2013a). Die Gewebequelle ist reichlich vorhanden und leicht zugänglich. Ethisch-medizinische Regularien kommen dabei nicht in Betracht. Anders als bei invasiver Entnahme von autologem Zellgewebe (Knochenmark, Fettgewebe) entfällt das Infektionsrisiko. Entzündliche Reaktionen oder immunologische Abwehrreaktionen sind nicht bekannt. Allerdings stützt sich der beobachtete Behandlungserfolg  $2GFR = 94,1 \%$  von Amniongewebe-MSK auf nur 1 Studie und bedarf daher der Bestätigung. Bemerkenswert ist, dass Fettgewebe-MSK (RICH et al. 2014) beim FT-Ursprung, wofür Vergleichsgruppen noch fehlen, eine ähnlich hohe  $GFR = 86,6 \%$  (Prüfzeit 13-92 M) aufweisen wie Amniongewebe-MSK (Tab 5.4.4.3).

### 6.3.2 KtrB

Bei der OBS führt der Vergleich von KtrB mit PRP und Knochenmark-MSK zu replizierbaren Ergebnissen von niedriger interner Evidenz. Die eingangs gestellte Frage, ob die Behandlung natürlich entstandener Erkrankungen des Sehnenapparates oder FT bei Sportpferden mit prT kombiniert mit kontrollierter Bewegung tatsächlich einer Behandlung ausschließlich mit kontrollierter Bewegung überlegen ist, lässt sich aufgrund der mangelnden Fallzahl und der kleinen Effektdifferenz ( $d_{GFR} = 2,9 \%$  bzw.  $11,3 \%$ ) zwischen KtrB und den beiden prT mittels Power-Analyse noch nicht klären (Tab 5.4.4.3).

### 6.3.3 PRP

Für PRP bei der OBS ergibt der Vergleich mittels Power-Analyse sowohl mit Knochenmark-MSK bei der OBS ( $d_{GFR} = 8,4 \%$ ; aktuelle Power  $17,7 \%$ ;  $p = 0,0447$ ) als auch mit PRP beim FT-Schenkel ( $d_{GFR} = 3,3 \%$ ; aktuelle Power  $< 5 \%$ ;  $p < 0,05$ ) ein replizierbares Ergebnis von niedriger interner Evidenz. Ob sich mit PRP die GF in beiden Fällen unterscheidet, kann aufgrund der mangelnden Fallzahl und kleinen Effektdifferenz noch nicht geklärt werden. In Verbindung mit der Frage, inwieweit sich physiologische Gewebeunterschiede zwischen OBS und FT-Schenkel auf das regenerative Potenzial von PRP auswirken, ist der Vergleich jedoch für zukünftige Studien interessant (Tab 5.4.4.3).

Sarah E. M. Doll

**Metaanalyse klinischer Studien 1983-2016 zur langfristigen Gebrauchsfähigkeit von Sportpferden nach Behandlung von natürlich entstandenen Erkrankungen der oberflächlichen und der tiefen Beugesehne und des Fesselträgers entweder allein mit kontrollierter Bewegung oder kombiniert mit einem potenziell regenerativen Therapeutikum**

Veterinärmedizinische Fakultät, Universität Leipzig, Klinik für Pferde.

Eingereicht 2019. 92 Seiten, 4 Abbildungen, 25 Tabellen, 181 Literaturangaben.

**Schlüsselwörter:** Metaanalyse, Pferd, natürlich entstandene Tendinopathie, Fesselträger, regenerativ, konservativ

**Einleitung:** Seit ca. 15 Jahren versucht die equine Regenerationsmedizin nach einer Sehnen- oder Bandverletzung die konsekutive Narbenbildung zu reduzieren und mit intraläsional applizierten potenziell regenerativen Therapeutika (prT) das ursprüngliche Gewebe wieder herzustellen. PrT sind mesenchymale stromale multipotente Progenitorzellen (MSC) aus unterschiedlichen Gewebequellen, sowie blutzellbasierende (meist PRP, Platelet Rich Plasma) oder azelluläre Blutprodukte. In 2001-2016 werden 19 verschiedene prT appliziert, die mithilfe unterschiedlicher labortechnischer Verfahren und durch Kombination aus 9 Gewebequellen entstehen: 3 *autologe* [Knochenmark (Bone Marrow BM); Fettgewebe (Adipose Ad); Blut], 4 *allogene* [BM; Ad; Blut; Amniongewebe (AM)], 2 *xenogene* [ovines AM, porcine Harnblasenmatrix]. Jedoch sind die vielversprechenden Ergebnisse der Studien nicht vergleichbar und fehlt bisher ein evidenter Nachweis des langfristigen Behandlungserfolgs. Um dies systematisch zu klären kommt nur die metaanalytische Methode in Frage und ist das Ziel dieser Dissertation.

**Ziele:** Wie kann mit einer einzigen Effektvariable die langfristige Gebrauchsfähigkeit von Sportpferden, die nach einer natürlich entstandenen Erkrankung der OBS, TBS oder FT konservativ (KtrB) oder regenerativ (PrTB) behandelt werden, quantitativ valide gemessen werden? Und wie können bisherige Studienergebnisse verglichen werden? Hypothesen: Es sind zwei Arbeitsschritte erforderlich. (1) Ein evidenzbasiertes einfaches Evaluationsmodells mit der neuen quantitativen Effektgröße 'Gebrauchsfähigkeitsrate' *GFR* kann entworfen werden. (2) Die metaanalytische 'gepoolte Reanalyse' kann mit diesem Modell die Ergebnisse der bisherigen Studien unabhängig von den Ergebnissen der Primärstudien neu bewerten und vergleichen.

**Material und Methoden:** Inkludiert sind die in der Arbeitsfrage genannten in vivo Studien. Exkludiert sind in vivo Studien mit künstlich induzierten Erkrankungen oder mit konventionellen oder chirurgischen Behandlungsmethoden und in vitro Studien. Untersucht werden 51 Studien in 1983-2016 (darunter 3 kleine RCTs); davon 30 (der 51) in 2011-2016; zusätzlich 21 überarbeitete Studien-Ergebnisse in 1983-2011 aus dem metaanalytischen Teil einer älteren Studie. Folgende in der Forschungsliteratur diskutierte Einflussfaktoren auf GF werden als Primärdaten definiert: PrT, Gewebestruktur, sportliche Disziplin, Läsionsgrad, Altersklasse der Pferde, Prüfzeitdauer (P). In P wird kontrolliert, wie viel der Sportpferde, die das Erfolgsmerkmal (EM) erreichen, ein Rezidiv erleiden. EM ist: 'Nach Rehabilitation kehren die Sportpferde zur Arbeit auf (mindestens) dem ursprünglichen Leistungsniveau zurück'. Die Daten werden aus den Studien extrahiert und in Behandlungsgruppen (BehG) nach Gewebestruktur (OBS, TBS, USB der TBS, FT-Ursprung/-Körper/-Schenkel) und Behandlungsform (KtrB, PrTB) geordnet. Da *GFR* in jeder BehG unter gleichen Bedingungen gemessen werden soll, wird im Konsens mit der Forschung die Studiendauer standardisiert. Die Studiendauer summiert sich aus der projizierten Rehabilitationsdauer (proj. Reha) und P; daher sind beide zu standardisieren. Die Berechnung von *GFR* aus nur 3 Variablen (A, B, C) wird normiert um *GFR* zwischen BehG vergleichbar zu machen. Kleine inhaltlich homogene und auf Bias geprüfte BehG werden mittels quantitativer Heterogenitätsanalyse zu größeren BehG mit einem präziseren Effektschätzer  $GFR_{\text{gepoolt}}$  (mit 95 %-CI) gepoolt. Die Power-Analyse kann interessierende Unterschiede zwischen zwei *GFR* per Signifikanznachweis statistisch sichern, wenn die erforderliche Fallzahl gegeben ist.

### 1) Ergebnis zum Entwurf des *GFR*-Modells:

*GFR* quantifiziert valide die GF von Sportpferden nach Behandlung mit PrTB oder KtrB für 1,5 Jahre (*a*) rezidivfreie Prüfzeit beim FT ( $1,5GFR$ ) bzw. für 2*a* bei OBS oder TBS ( $2GFR$ ). Die standardisierte Studiendauer summiert sich bei OBS, TBS auf insgesamt 3*a* (= 1*a* proj. Reha + 2*a* P); beim FT auf 2,25*a* (= 0,75*a* proj. Reha + 1,5*a* P). Die Zeit, die ein Patient individuell zur Rehabilitation benötigt, ist im Modell nicht limitiert. Die Berechnung von *GFR* ist normiert durch  $GFR = (N-A-B-C) / (N-A)$ ;  $GFR = E(1-R)$ ; Erfolgsrate (bzgl. EM) ist  $E = (N-A-B-C) / (N-A)$ ; Rezidivrate ist  $R = C / (N-A-B)$ . A, B, C repräsentieren disjunkte Fallgruppen, somit ist jeder Patient genau einer zugeordnet; ihre Größe kann am Ende der Studie leicht abgezählt werden. N ist die Gesamtzahl der behandelten Patienten einer BehG. A Patienten scheiden aus Gründen aus, die mit dem Behandlungsmittel kausal nicht zusammenhängen (z.B. Tod, Zuchtzwecke, Besitzerwille) oder werden nicht regelkonform behandelt (z. B. das Bewegungsprogramm wird nicht befolgt); diese sind bei der Berechnung von *GFR* zu eliminieren um einen Bias zu vermeiden. N-A = n\* Patienten werden somit regelkonform behandelt. B Patienten bleiben allein wegen der mangelnden Wirkung des Behandlungsmittels (prT) rekonvaleszent. N-A-B Patienten erreichen das EM und werden während P auf ein Rezidiv kontrolliert; C von N-A-B erleiden ein Rezidiv.

### 2) Ergebnisse zur Heterogenitätsanalyse:

Aus den 51 Studien lassen sich 100 BehG (69 in 2011-2016; 31 in 2001-2011) extrahieren. Bei 50 (der 100) BehG kann *GFR* (incl. 95 %-CI) bestimmt werden (45 in 2011-2016; 5 in 2001-2011). 13 (der 50) BehG haben die Fallzahl  $30 \leq n^* \leq 145$ ; 6 (der 13) BehG differenzieren die Gewebestruktur. Durch Poolen entstehen 1 BehG (aus 3 BehG) zu KtrB ( $2GFR = 57,5 \%$ ,  $n^* = 60$ , OBS) und 3 BehG (aus je 2 BehG) zu PrTB: BM-MSc ( $2GFR = 68,8 \%$ ,  $n^* = 145$ , OBS); PRP ( $2GFR = 60,4 \%$ ,  $n^* = 51$ , OBS); PRP ( $GFR = 63,7 \%$ ,  $n^* = 42$ , FT-Schenkel). Hinzu kommen 2 einzelne BehG zu PrTB: Ad-MSc ( $GFR = 86,6 \%$ ,  $n^* = 30$ , FT-Ursprung); AM-MSc ( $2GFR = 94,1 \%$ ,  $n^* = 34$ , OBS). Insgesamt kann erstmals mittels Power-Analyse GF anhand adäquater BehG-Paare, die aus diesen 6 BehG kombiniert werden, qualitativ und quantitativ valide verglichen werden.

### 3) Ergebnisse zur Power-Analyse:

(1) **Verschiedene Behandlungsformen bei der OBS.** Die Chance OR (Odds Ratio) für die langfristige Gebrauchsfähigkeit ( $2GFR$ ) der Sportpferde mit AM-MSc erscheint mindestens 7-mal so groß sowohl vs. BM-MSc ( $OR = 7,2$  [1,7-31,4]) als auch vs. PRP ( $OR = 9,7$  [2,1-45]) sowie vs. KtrB ( $OR = 11,4$  [2,5-52,1]). Erstmals kann mittels Power-Analyse die Effektdifferenz  $d_{GFR} > 25 \%$  zwischen diesen BehG per Signifikanznachweis (aktuelle Power  $> 90 \%$ ; aktuelles  $p < 0,025$ ) durch die erforderliche Fallzahl statistisch gesichert werden. Allerdings stützt sich  $2GFR$  für AM-MSc auf nur eine Studie. Dagegen fehlt bei  $d_{GFR} < 25 \%$  zwischen PRP vs. BM-MSc, KtrB vs. BM-MSc, KtrB vs. PRP die erforderliche Fallzahl für den entsprechenden Nachweis.

(2) **Gleiche Behandlungsform bei verschiedenen Gewebestrukturen.** Ob ein Unterschied zwischen GF mit PRP bei der OBS und beim FT-Schenkel besteht, bleibt aufgrund mangelnder erforderlicher Fallzahl und der sehr kleinen Effektdifferenz  $d_{GFR} = 3,3 \%$  ungeklärt.

(3) **Gewebestruktur und Behandlungsform sind verschieden.** Vergleiche sind nicht möglich. Jedoch ist bemerkenswert:  $2GFR = 86,6 \%$  mit Ad-MSc beim FT-Ursprung scheint ähnlich hoch wie  $2GFR = 94,1 \%$  mit AM-MSc bei der OBS.

**Schlussfolgerung:** Beim gegenwärtigen Stand der Forschung erscheint bei der OBS die langfristige Gebrauchsfähigkeit ( $2GFR$ ) mit AM-MSc gegenüber BM-MSc, PRP sowie KtrB überlegen. Beim FT fehlen Vergleichsgruppen noch. Eine größere Fallzahl ist erforderlich um: (1) die Power-Analyse auch für kleine Effektdifferenzen  $d_{GFR} < 25 \%$  zu ermöglichen, (2) den Einfluss auf *GFR* durch Läsionsgrad, Altersklasse oder Herstellungsverfahren für prT zu klären, (3) die Verkürzung der Rehabilitationsdauer nachzuweisen. Studien gemäß *GFR*-Modell können replizierbare Ergebnisse von hoher interner Evidenz erzielen. Vergleichbar dem

Konzept prospektiv geplanter Metaanalysen bieten zukünftige Studien die Chance auf Ergebnisse von hoher externer Evidenz.

## 8 SUMMARY

Author: Sarah E. M. Doll

**Meta-analysis of clinical studies in 1983-2016 for 'long-term fit for use' (FfU) of sport horses after treatment of naturally occurring tendinopathies of the superficial (SDFT), deep digital flexor tendon (DDFT) or suspensory ligament (SL) either with 'controlled exercise alone' (CtrT) or 'combined with potential regenerative therapeutics' (PrTT)**

University Equine Hospital, Faculty of Veterinary Medicine, University of Leipzig. Submitted in 2019. 92 pages, 4 figures, 25 tables, 181 references.

**Keywords:** meta-analysis, naturally occurring tendinopathy, suspensory ligament, regenerative, conservative

**Introduction:** About 15 years equine regenerative medicine is seeking to reduce the subsequent scarring after a tendon or ligament injury and to restore the original tissue with intralesionally applicated 'potential regenerative therapeutics' (**prT**). PrT are mesenchymal stromal multipotential progenitor cells (**MSC**) from different sources of tissue, as well as products of blood, which are based on blood cells (e. g. **PRP**, platelet rich plasma) or are acellular. In 2001-2016 19 different prT were administered, produced by various techniques and combinations of 9 sources of tissue: 3 autologous [bone marrow (BM); adipose (Ad); blood], 4 allogeneous [bone marrow, adipose, blood, amnion (AM)], 2 xenogeneous [ovine amnion; porcine urinary bladder matrix]. However the promising results of research studies are not comparable and evidence of long-term treatment-success has not been provided to date. To clarify this systematically only a meta-analytical method is plausible and is the aim of this dissertation.

**Objectives:** How can the 'long-term fit for use' (FfU) of sport horses, which are treated conservatively (CtrT) or regeneratively (PrTT) after naturally occurring tendinopathies of SDFT, DDFT or FT, be quantified in a valid way with a single effect variable? And how can the results of previous research studies be compared?

**Hypotheses:** Two sequential working steps are required. (1) An evidence-based simple evaluation-model with the new quantitative effect size called 'rate of fit for use' **FFU** can be developed. (2) Based on this model a 'pooled re-analysis' can re-assess and compare results of previous research studies independently from results of primary studies.

**Material and Methods:** Included are in vivo studies as defined in the leading question. Excluded are in vivo studies with artificially induced lesions or conventional or surgical treatment and in vitro studies. 51 research studies conducted in 1983-2016 (3 of it are small RCTs) are considered, 30 (of 51) in 2011-2016; in addition, 21 revised findings of studies conducted in 1983-2011 out of the meta-analytic section of an earlier study. The following influencing factors affecting FfU discussed in the research literature are defined as primary data: PrT, structure of tissue, sportive kind-of-use, degree of lesion, age class of horses, duration of monitoring (**DoM**). During DoM it is controlled how many of the sport horses, which perform the outcome (**OC**), re-injure the tissue. OC is: 'After Rehabilitation the sport horses return to work at (not less than) the initial level of performance'.

The data are extracted from the studies and configured to treatment groups (**TGs**) according to the structure of tissue (SDFT, DDFT, SL-body, SL-origin, SL-branches) and the kind of treatment (CtrT, PrTT). As **FFU** has to be measured for each TG under the same conditions, the duration of the study is standardized in consensus with research studies. The duration of the study is summarized from the planned duration of rehabili-

tation (**PlanReha**) and DoM; so both need to be standardized. *FFU* is calculated from just 3 variables (A, B, C) according to norm formula to make *FFU* comparable between TGs. Small TGs, homogen in terms of content and checked for bias, are pooled by a quantitative analysis of heterogeneity to a more increased TG with more precise estimated treatment effect  $FFU_{\text{pooled}}$  (incl. 95 %-CI). Power-analysis can statistically assure interesting differences between two *FFU* by significant evidence, if there is given the needed number of cases.

1) **Results of *FFU*-model:** *FFU* quantifies in a valid way the FfU of sport horses after treatment with CtrT or PrTT remaining free of recurrence for 1.5 years (*a*) DoM at SL ( $1.5FFU$ ) respectively 2 *a* DoM at SDFT or DDFT ( $2FFU$ ). The standardized duration of study sums up at SL to 2.25 *a* (= 0.75 *a* PlanReha + 1.5 *a* DoM); respectively at SDFT or DDFT to 3 *a* (= 1 *a* PlanReha + 2 *a* DoM). The individual duration of rehabilitation is not limited under the model. *FFU* is calculated according to the norm formula  $FFU = (N-A-B-C) / (N-A)$ ; respectively is  $FFU = E(1-R)$ ; 'rate of OC' is  $E = (N-A-B-C) / (N-A)$ ; 'rate of recurrence' is  $R = C / (N-A-B)$ . A, B, C represent disjunctive case groups of N, so that each patient can be assigned to just one group. They can easily be counted having finished the practical phase of study. N is the total number of treated patients of a TG. A patients are retired by reasons causally unrelated to treatment agent (e. g. death, breeding purpose, owner willingness) or don't comply with the study regulations (e. g. the controlled exercise program is not obeyed). They need to be eliminated when calculating *FFU* in order to prevent bias. Consequently  $N-A = n^*$  patients are treated complying with regulations. B patients remain reconvalescent because of inefficacy of the treatment agent (prT) solely. In sum, N-A-B patients perform OC and are controlled for re-injury during DoM; C of N-A-B suffer a recurrence.

2) **Results of analysis of heterogeneity:** From the 51 studies 100 TGs can be extracted (69 in 2011-2016; 31 in 2001-2011). *FFU* (incl. 95 %-CI) can be determined with 50 (of 100) TGs (45 in 2011-2016; 5 in 2001-2011). 13 (of 50) TGs have the case number  $30 \leq n^* \leq 145$ ; 6 (of 13) TGs differentiate the structure of tissue. By pooling 1 TG (from 3 TGs) at CtrT ( $2GFR = 57.5\%$ ,  $n^* = 60$ , SDFT) and 3 TGs (each from 2 TGs) at PrTT: BM-MSc ( $2FFU = 68.8\%$ ,  $n^* = 145$ , SDFT); PRP ( $2FFU = 60.4\%$ ,  $n^* = 51$ , SDFT); PRP ( $FFU = 63.7\%$ ,  $n^* = 42$ , LS-branches) are obtained. In addition, there are 2 single TGs at PrTT: Ad-MSc ( $2FFU = 86.6\%$ ,  $n^* = 30$ , LS-origin); AM-MSc ( $2FFU = 94.1\%$ ,  $n^* = 34$ , SDFT). Overall, for the first time by power-analysis FfU can be compared with adequate TG-pairs, which are combined by these 6 TGs, qualitatively and quantitatively in a valid way.

### 3) Results of power-analysis:

(1) **Different treatments at SDFT:** The chance *OR* (odds ratio) for the long-term 'fit for use' ( $2FFU$ ) of sport horses with AM-MSc appears to be more than at least 7-fold as high as BM-MSc ( $OR = 7.2 [1.7-31.4]$ ), 9-fold as PRP ( $OR = 9.7 [2.1-45]$ ) and 11-fold as CtrT ( $OR = 11.4 [2.5-52.1]$ ). For the first time, the power-analysis can statistically assure the difference of effect  $25\% < d_{FFU}$  (current power  $> 90\%$ ; current  $p < 0.025$ ) between these TGs through significant evidence based on the needed number of cases. However  $2FFU$  of AM-MSc relies on only 1 study. By contrast the needed number of cases is insufficient to statistically assure the difference of effect  $d_{FFU} < 25\%$  by significant evidence between PRP vs. BM-MSc, CtrT vs. BM-MSc, CtrT vs. PRP. (2) **Same treatment at different structures of tissue:** It remains unclear whether there is a difference between the long-t. FfU with PRP at SDFT or SL-branches because the number of cases is insufficient and the difference of effect  $d_{FFU} = 3.3\%$  is very small. (3) **Tissue structure and treatment are different:** No comparisons are possible. Noticeable is:  $2FFU = 86.6\%$  with Ad-MSc at SL-origin seems as high as  $2FFU = 94.1\%$  with AM-MSc at SDFT.

**Conclusion:** At the present stage of research the *FFU* ( $2FFU$ ) at SDFT with AM-MSc appears to be superior to BM-MSc as well as to PRP and to CtrT. At SL comparison groups still are lacking. A greater number of cases will be required: (1) to perform power-analysis also for small differences of effect  $d_{FFU} < 25\%$ , (2) to explore impact on *FFU* of grade of lesion, of age class or of manufacturing process for prT, (3) to verify the

evidentiary reduction of duration of rehabilitation. Studies under *FFU*-model can obtain replicable results of high internal evidence; according to the concept of prospective planned meta-analysis studies in the future offer the chance of results of high external evidence.



Abellanet de Oleza I. Treatment of damaged soft tissue and joints using Platelet-rich-plasma in race horses: clinical and biochemical evidence supporting its use. (Originaltitel in spanisch) [Dissertation med. vet.]. Barcelona: Univ. Barcelona; 2009.

Arlt S. Untersuchungen zur Einbindung der evidenzbasierten Reproduktionsmedizin in die veterinärmedizinische Lehre und Praxis. [Habilschr. med. vet.] Berlin: Freie Univ. Berlin; 2013.

Avella CS, Ely ER, Verheyen KL, Price JS, Wood JL, Smith RK. Ultrasonographic assessment of the superficial digital flexor tendons of National Hunt racehorses in training over two racing seasons. *Equine Vet J.* 2009;41:449-54.

Banse M. Der Einfluss von regenerativen Therapien auf die Heilung des Musculus interosseus medius beim Pferd: Eine kontrollierte prospektive klinische und sonographische Studie. [Dissertation med. vet.]. Gießen: Univ. Gießen; 2011.

Bazzano M, Piccione G, Giannetto C, Tosto F, Di Pietro S, Giudice E. Platelet Rich Plasma Intralesional Injection as Bedside Therapy for Tendinitis in Athletic Horse. *Acta Sci Vet.* 2013;41:145.

Beerts C, Seifert C, Zimmerman M, Felix E, Suls M, Mariën T, Broeckx S, Spaas JH. Desmitis of the Accessory Ligament of the Equine Deep Digital Flexor Tendon: A Regenerative Approach. *J Tissue Sci Eng.* 2013; 4:1.

Blaber SP, Webster RA, Hill CJ, Breen EJ, Kuah D, Vesey G, Herbert BR. Analysis of in vitro secretion profiles from adipose-derived cell populations. *J Transl Med.* 2012;10:172.

Blettner M, Sauerbrei W, Schlehofer B, Scheuchenpflug T, Friedenreich C. Vergleich von traditionellen Reviews, Metaanalysen und gepoolten Analysen zur Bewertung von Risikofaktoren. *Informatik, Biometrie und Epidemiologie in Medizin und Biologie.* 1997;28 (3):148-66.

Blettner M, Sauerbrei W. Metaanalysen epidemiologischer Studien. *Medizinische Klinik.* 1998;93:442-5.

Blobel K. Ein Beitrag zur Sehnenbehandlung bei Reit- und Trabrennpferden. *Prakt Tierarzt.* 1988;69:10-14.

Bosch G, Moleman M, Barneveld A, van Weeren PR, van Schie HTM. The effect of platelet-rich plasma on the neovascularization of surgically created equine superficial digital flexor tendon lesions. *Scand J Med Sports.* 2011;21:554-61.

Brehm W, Burk J, Dellinh U, Gittel C, Ribitsch I. Stem cell-based tissue engineering in veterinary orthopedics. *Cell Tissue Res.* 2012;347:677-88.

Brossi PM, Moreira JJ, Machado TS, Baccarin RY. Platelet-rich plasma in orthopedic therapy: a comparative systematic review of clinical and experimental data in equine and human musculoskeletal lesions. *BMC Vet Res.* 2015;11:98.

Burk J. Klinische Anwendung und vergleichende Charakterisierung equiner mesenchymaler Stromazellen [Dissertation med. vet.]. Leipzig: Uni Leipzig; 2012.

Burk J, Brehm W. Stammzellentherapie von Sehnenverletzungen – klinische Ergebnisse von 98 Fällen. *Pferdeheilkd.* 2011;27:153-61.

Burk J, Troillet A. Regenerative Therapieformen. In: Brehm W, Gehlen H, Ohnesorge B, Wehrend A, Hrsg. *Handbuch Pferdepraxis.* 4. Aufl. Stuttgart: Enke; 2016. S. 882-91.

- Butcher MT, Hermanson JW, Ducharme NG, Mitchell LM, Soderholm LV, Bertram JE. Superficial digital flexor tendon lesions in racehorses as a sequella to muscle fatigue: a preliminary study. *Equine Vet J.* 2007; 39:540-54.
- Butler DL, Juncosa-Melvin N, Boivin GP, Galloway MT, Shearn JT, Gooch C, Awad H. Functional tissue engineering for tendon repair: a multi disciplinary strategy using mesenchymal stem cells, bioscaffolds and mechanical stimulation. *J Orthop Res.* 2008;26:1-9.
- Calzetta I, Roncada P, di Cave D, Bonizzi L, Urbani L, Pistocchini E, Rogliani P, Matera MG. Pharmacological treatments in asthma-affected horses: A pair-wise and network meta-analysis. *Equine Vet J.* 2017;49:710-17.
- Caplan AI. Mesenchymal stem cells. *J Orthop Res.* 1991;9:641-50.
- Caplan AI. Mesenchymal stem cells as trophic mediators. *J Cell Biochem.* 2006;98:1076-84.
- Caplan AI. Why are MSCs therapeutic? New data: new insight. *J Pathol.* 2009;217:318-24.
- Castelijns G, Crawford A, Schaffer J, Ortolano GA, Beauregard T, Smith RK. Evaluation of a filter-prepared platelet concentrate for the treatment of suspensory branch injuries in horses. *Vet Comp Orthop Traumatol.* 2011;24:363-69.
- Chalmers I, Hedges LV, Cooper H. A brief history of research synthesis. *Eval Health Prof.* 2002;25:12-37.
- Chesen AB, Dabareiner RM, Chaffin MK, Carter GK. Tendinitis of the proximal aspect of the superficial digital flexor tendon in horses: 12 cases. *J Am Vet Med Assoc.* 2009;234:1432-36.
- Cohen J. *Statistical Power Analysis for the Behavioral Sciences.* 1988 (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Coudry V, Dupays AG, Carnicer D, Jacquet S, Bertoni L, Crevier-Denoix N, Desquilbet L, Kichenin K, Barritault D, Denoix JM. Long-Term Follow-up of Superficial Digital Flexor Tendonitis Treated by a Single Intralesional Injection of a ReGeneraTing Agent in 51 Horses. *J Equine Vet Sci.* 2014;34:1357-60.
- Cowles RR. Proximal Suspensory Desmitis – A Qualitative Survey. *Proceedings of the Annual Convention of the American Association of Equine Practitioners. Am Ass Equine Practnrs.* 2000;46:143-44.
- Crevier-Denoix N, Robin D, Pourcelot P, Falala S, Holden L, Estoup P, Desquilbet L, Denoix J, Chateau H. Ground reaction force and kinematic analysis of limb loading on two different beach sand tracks in harness trotters. *Equine Vet J Suppl.* 2010;38:544-51.
- Crovace A, Lacitignola L, Rossi G, Francioso E. Histological and immunohistochemical evaluation of autologous cultured bone marrow mesenchymal stem cells and bone marrow mononucleated cells in collagenase-induced tendinitis of equine superficial digital flexor tendon. *Vet Med Int.* 2010.
- Dahlgren LA, Mohammed HO, Nixon AJ. Temporal expression of growth factors and matrix molecules in healing tendon lesions. *J Orthop Res.* 2005;23:84-92.
- Deeks JJ, Higgins JPT, Altman DG. Analysing data and undertaking meta-analyses. In: Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011].* The Cochrane Collaboration, 2011. Available from <http://handbook.cochrane.org>.
- Dahlgren LA. Pathobiology of tendon and ligament injuries. *Clin Tech Equine Pract.* 2007;6:168-73.
- De Schauwer C, Van de Walle GR, Van Soom A, Meyer E. Mesenchymal stem cell therapy in horses: useful beyond orthopedic injuries? *Vet Q.* 2013;33:234-41.

Docheva D, Müller SA, Majewski M, Evans CH. Biologics for tendon repair. *Adv Drug Deliv Rev.* 2015;84:222-39.

Dominici M, Le Blanc K, Müller I, Slaper-Cortenbach I, Marini FC, Krause DS, Deans RJ, Keating A, Prockop DJ, Horwitz EM. Minimal criteria for defining multipotent mesenchymal stromal cells. The International Society for Cellular Therapy position statement. *Cytotherapy* 2006;8:315-17.

Dyson SJ. Treatment of superficial digital flexor tendonitis: A comparison of conservative management, sodium hyaluronate, and glycosaminoglycan polysulfate. *Proceedings of the 43rd Annual Convention of the American Association of Equine Practitioners; 1997 Dec 7-10; Phoenix, Arizona, USA. Lexington, KY. Am Ass Equine Practnrs.* 1997;43:297-300.

Dyson SJ. Proximal Suspensory Desmitis in the Forelimb and the Hindlimb. *Proceedings of the 46th Annual Convention of the American Association of Equine Practitioners; San Antonio, Texas, USA, Nov 26-29. Lexington, KY. Am Ass Equine Practnrs.* 2000;46:137-142.

Dyson SJ, Genovese RL. The suspensory apparatus. In: Ross MW, Dyson SJ, Hrsg. *Diagnosis and Management of Lameness in the Horse.* 1. Aufl. Philadelphia: Saunders; 2003. p. 654-666.

Dyson SJ. Medical management of superficial digital flexor tendonitis: a comparative study in 219 horses (1992–2000). *Equine vet J.* 2004;36:415-19.

Dyson SJ. Diagnosis and management of common suspensory lesions in the forelimbs and hindlimbs of sport horses. *Clin Tech equine Pract.* 2007;6:177-88.

Edinger J. Orthopädische Untersuchung der Gliedmaßen und der Wirbelsäule. In: Wissdorf H, Gerhards H, Huskamp B, Deegen E, Hrsg. *Praxisorientierte Anatomie und Propädeutik des Pferdes.* 3. Aufl. Hannover: Schaper; 2010. p. 890-926.

Freedman KB, Back S, Bernstein J. Sample size and statistical power of randomized, controlled trials in orthopaedics. *J Bone Joint Surg (Br).* 2001;19:528-35.

Frisbie DD, Smith RK. Clinical update on the use of mesenchymal stem cells in equine orthopaedics. *Equine Vet J.* 2010;42:86-9.

Garrett KS, Bramlage LR, Spike-Pierce DL, Cohen ND. Injection of Platelet- and Leukocyte-Rich Plasma at the Junction of the Proximal Sesamoid Bone and the Suspensory Ligament Branch for Treatment of Yearling Thoroughbreds with Proximal Sesamoid Bone Inflammation and Associated Suspensory Ligament Branch Desmitis. *J Am Vet Med Assoc.* 2013;243:120-25.

Geburek F, Stadler PM. Regenerative Therapie von Sehnen- und Bänderkrankungen bei Pferden. Terminologie, Herstellung, biologisches Potenzial und In-vitro-Effekte. *Tierärztliche Praxis Großtiere.* 2011a;39:373-83.

Geburek F, Stadler PM. Regenerative Therapie von Sehnen- und Bänderkrankungen bei Pferden: Ergebnisse der Behandlung mit Stammzellen, Blutprodukten, Gerüstsubstanzen und Wachstumsfaktoren – eine Literaturübersicht und Metaanalyse. *Pferdeheilkd.* 2011b;27:609-25.

Geburek F, Lietzau M, Beineke A, Rohn K, Stadler PM. Effect of a single injection of autologous conditioned serum (ACS) on tendon healing in equine naturally occurring tendinopathies. *Stem Cell Res Ther.* 2015;26:126 (doi:10.1186/s13287-015-0115-0).

Geburek F, Gaus M, van Schie HT, Rohn K, Stadler PM. Effect of intralesional platelet-rich plasma (PRP) treatment on clinical and ultrasonographic parameters in equine naturally occurring superficial digital flexor tendinopathies – a randomized prospective controlled clinical trial. *BMC Vet Res.* 2016;12:191.

Geburek F, Roggel F, van Schie HT, Beineke A, Estrada R, Weber K, Hellige M, Rohn K, Jagodzinski M, Welke B, Hurschler C, Conrad S, Skutella T, van de Lest C, van Weeren R, Stadler PM. Effect of single intra-lesional treatment of surgically induced equine superficial digital flexor tendon core lesions with adipose-derived mesenchymal stromal cells: a controlled experimental trial. *Stem Cell Res Ther* 2017;8:129 (doi:10.1186/s13287-017-0564-8).

Genovese RL, Rantanen NW, Hauser M, Simpson B. Diagnostic ultrasonography of equine limbs. *Vet Clin North Am Equine Pract.* 1986;2:145-226.

Genovese RL, Longo K, Berthold B, Jorgensen JL. Quantitative sonographic assessment in the clinical management of superficial digital flexor injuries in thoroughbred racehorses. *Proceedings of the 43rd Annual Convention of the American Association of Equine Practitioners*; 1997 Dec 7-10; Phoenix, Arizona, USA. Lexington, KY. *Proc Am Ass Equine Practnrs.* 1997;43:285-90.

Gibson KT, Burbidge HM, Pfeiffer DU. Superficial digital flexor tendonitis in Thoroughbred race horses: outcome following non-surgical treatment and superior check desmotomy. *Aust Vet J.* 1997;75:631-35.

Giese M. Die besondere Biologie der Stammzellen – ein Review. LBH: 8. Leipziger Tierärztekongress – Tagungsband 2; Leipzig 2016. S. 12.

Gillis C. Rehabilitation of tendon and ligament injuries. *Proceedings of the 43rd Annual Convention of the American Association of Equine Practitioners*; 1997 Dec 7-10; Phoenix, Arizona. Lexington, KY. *Am Ass Equine Practnrs.* 1997;43:306-9.

Glass GV. Primary, Secondary and Meta-Analysis of Research. *Educ Researcher.* 1976;5:3-8.

Goletz I. Isolierung und Charakterisierung equiner mesenchymaler Stammzellen für einen möglichen Einsatz im Tissue Engineering [Dissertation med. vet.]. Gießen: Univ. Gießen; 2009.

Goodship AE, Birch HL, Wilson AM. The pathobiology and repair of tendon and ligament injury. *Vet Clin North Am Equine Pract.* 1994;10:323-49.

Godwin EE, Young NJ, Dudhia J, Beamish IC, Smith RK. Implantation of bone marrow-derived mesenchymal stem cells demonstrates improved outcome in horses with overstrain injury of the superficial digital flexor tendon. *Equine Vet J.* 2012;44:25-32.

Guercio A, Di Marco P, Casella S, Russotto L, Puglisi F, Majolino C, Giudice E, Di Bella S, Purpari G, Cannella V, Piccione G. Mesenchymal Stem Cells Derived From Subcutaneous Fat and Platelet-Rich Plasma Used in Athletic Horses With Lameness of the Superficial Digital Flexor Tendon. *J Equine Vet Sci.* 2015;35:19-26.

Hall MS, Vasey JR, Russell JW, Russell T. Use of ultrasound-guided autologous bone marrow transfer for treatment of suspensory ligament desmitis in 30 race horses (2003-2010). *Aust Vet J.* 2013;91:102-7.

Hansen HF, Rieper O. The evidence movement: The development and consequences of methodologies in review practices. *Evaluation.* 2009;15:141-63.

Heintel KE. Untersuchungen zu biomechanischen Eigenschaften von Gleit- und Zugsehnern [Dissertation med. vet.]. München: LMU München; 2013.

Higgins JP. Commentary: heterogeneity in meta-analysis should be expected and appropriately quantified. *Int J Epidemiol* 2008;37:1158-60.

Higgins JP, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from <http://handbook.cochrane.org>.

- Higgins JP, Thompson SG. Quantifying heterogeneity in a metaanalysis. *Stat Med* 2002;21:1539–58.
- Herthel DJ. Enhanced Suspensory Ligament Healing in 100 Horses by Stem Cells and other Bone Marrow Components. Proceedings of the 47th Annual Convention of the American Association of Equine Practitioners; 2001 Nov 24-28; San Diego, California, USA. Lexington, KY. *Am Ass Equine Practnrs.* 2001;47:319-21.
- Hofberger S, Gauff F, Licka T. Suspensory ligament degeneration associated with pituitary pars intermedia dysfunction in horses. *Vet J.* 2015;203:348-50.
- Hu AJ, Bramlage LR. Racing performance of Thoroughbreds with superficial digital flexor tendonitis treated with desmotomy of the accessory ligament of the superficial digital flexor tendon: 332 cases (1989-2003). *JAVMA.* 2014;244:1441-48.
- Iacono E, Merlo B, Romagnoli M, Rossi B, Ricci F, Spadari A. Equine Bone Marrow and Adipose Tissue Mesenchymal Stem Cells: Cytofluorimetric characterization, in vitro differentiation and clinical application. *J Equine Vet Sci.* 2015;35:130-40.
- Johnston C, Gottlieb-Vedi M, Drevemo S, Roepstorff L. The kinematics of loading and fatigue in the standardbred trotter. *Equine Vet J Suppl.* 1999;30:249-53.
- Jorgensen JS, Genovese RL, Ross MW. Superficial digital flexor tendonitis in racehorses. In: Ross MW, Dyson SJ, Hrsg. *Diagnosis and Management of Lameness in the Horse.* 2. Aufl. St. Louis: Elsevier, Saunders; 2011. p. 706-15.
- Jüni P, Altman DG, Egger M. Assessing the quality of controlled clinical trials. *Brit Med J.* 2001;323:42-6.
- Kaminski M. Histologische Untersuchungen der Normalstruktur des Fesselträgers und der Beugesehnen beim Pferd [Dissertation med. vet.]. Zürich: Univ. Zürich; 2006.
- Kang JG, Park SB, Seo MS, Kim HS, Chae JS, Kang KS. Characterization and clinical application of mesenchymal stem cells from equine umbilical cord blood. *J Vet Sci.* 2013;4:367-71.
- Karanicolas PJ, Farrokhyar F, Bhandari M. Blinding: who, what, when, why, how? *Can J Surg.* 2010;53:345-8.
- Kasashima Y, Smith RKW, Birch HL, Takahashi T, Kusano K, Goodship AE. Exercise induced tendon hypertrophy: cross sectional area changes during growth are influenced by exercise. *Equine Vet J Suppl.* 2002;34:264-68.
- Kasashima Y, Takahashi T, Smith RK, Goodship AE, Kuwano A, Ueno T, Hirano S. Prevalence of superficial digital flexor tendonitis and suspensory desmitis in Japanese Thoroughbred flat racehorses in 1999. *Equine Vet J.* 2004;36:346-50.
- Koch TG, Berg LC, Betts DH. Current and future regenerative medicine – principles, concepts, and therapeutic use of stem cell therapy and tissue engineering in equine medicine. *Can Vet J.* 2009;50:155-65.
- Kojah K, Vogel M, Hagen J. Precision and accuracy of repeat ultrasound image acquisition and analysis of the cross-sectional areas of the equine flexor tendons of the forelimbs for follow-up assessments. *Pferdeheilkd.* 2017;33:320-28.
- Kriz J, Lisch R. *Methodenlexikon für Mediziner, Psychologen, Soziologen.* 1. Aufl. München: Psychologie Verlags Union; 1988.
- Künneke A, Jaugstetter H, Heyers P. Bone marrow-autologous conditioned plasma (BM-ACP) beim Pferd – erste klinische Erfahrungen einer neuen, autologen Therapieform bei der Behandlung von Sehnenerkrankungen des Pferdes. *Pferdeheilkd.* 2008;24:519-23.

- Lacitignola L, Crovace A, Rossi G, Francioso E. Cell therapy for tendinitis, experimental and clinical report. *Vet res commun*. 2008;32:33–8.
- Lange-Consiglio A, Tassan S, Corradetti B, Meucci A, Perego R, Bizzaro D, Cremonesi F. Investigating the efficacy of amnion-derived compared with bone marrow-derived mesenchymal stromal cells in equine tendon and ligament injuries. *Cytotherapy*. 2013 a;15:1011-20.
- Lange-Consiglio A, Rossi D, Tassan S, Perego R, Cremonesi F, Parolini O. Conditioned medium from horse amniotic membrane-derived multipotent progenitor cells: immunomodulatory activity in vitro and first clinical application in tendon and ligament injuries in vivo. *Stem Cells Dev*. 2013B;22:3015-24.
- Lean IJ, Rabiee AR, Duffield TF, Dohoo IR. Invited review: Use of meta-analysis in animal health and reproduction: methods and applications. *J Dairy Sci*. 2009;92:3545-65.
- Lempe RA. Vergleichende magnetresonanztomographische Untersuchung am plantaren Aspekt des Tarsus des Pferdes – sonographische und histologische Korrelation des M. interosseusmedius an der Hintergliedmaße des Pferdes [Dissertation med. vet.]. Leipzig: Univ. Leipzig; 2007.
- Leppänen M, Miettinen S, Mäkinen S, Wilpola P, Katiskalahti T, Heikkilä P, Tulamo RM. Management of equine tendon & ligament injuries with expanded autologous adipose derived mesenchymal stem cells: a clinical study. *Regen Med (Suppl. 2)*. 2009 a;4:21.
- Leppänen M, Heikkilä P, Katiskalahti T, Tulamo RM. Follow up of recovery of equine tendon & ligament injuries 18-24 months after treatment with enriched autologous adipose-derived mesenchymal stem cells: a clinical study. *Regen Med (Suppl. 2)*. 2009 b;4:21.
- Lischer CJ, Ringer SK, Schnewlin M, Imboden I, Fürst A, Stöckli M, Auer J. Treatment of chronic proximal suspensory desmitis in horses using focused electrohydraulic shockwave therapy. *Schweiz Arch Tierheilkd*. 2006;148:561–8.
- Lovati AB, Corradetti B, Lange Consiglio A, Recordati C, Bonacina D, Cremonesi F. Characterization and differentiation of equine tendon-derived progenitor cells. *J Biol Regul Homeost Agents (Suppl. 2)*. 2011;25:75-84.
- Lutter JD, Schneider RK, Sampson SN, Cary JA, Roberts GD, Vahl CI. Medical treatment of horses with deep digital flexor tendon injuries diagnosed with high-field-strength magnetic resonance imaging: 118 cases (2000-2010). *J Am Vet Med Assoc*. 2015;247:1309-18.
- Lutz SR. Intraläsionale Anwendung von autologem, thrombozytenangereicherten Plasma und xenogener, azellulärer porziner Matrix bei Fesselträgerläsionen beim Pferd – eine klinische vergleichende Studie [Dissertation med. vet.]. Leipzig: Univ. Leipzig; 2011.
- Malynicz G. Evidence-based medicine. *Vet Rec*. 1998;143:619.
- Marfe G, Rotta G, De Martino L, Tafani M, Fiorito F, Di Stefano C, Poletti M, Ranalli M, Russo MA, Gambacurta A. A new clinical approach: use of blood-derived stem cells (BDSCs) for superficial digital flexor tendon injuries in horses. *Life Sci*. 2012;90:825-30.
- Marneris D, Dyson SJ. Clinical features, diagnostic imaging findings and concurrent injuries in 71 sports horses with suspensory branch injuries. *Equine Vet Educ*. 2014;26:281-335 (doi.org/10.1111/eve.12175).
- Marr CM, Love S, Boyd JS, McKellar Q. Factors affecting the clinical outcome of injuries to the superficial digital flexor tendon in National Hunt and point-to-point racehorses. *Vet Rec*. 1993;132:476–9.

- Martin E, Nielsen S, Quinn JC. Willingness of a Convenience Sample of Horse Owners to Use Stem Cell Therapy for the Treatment of Equine Tendon Injury and Factors that Influence This Decision. *J Equine Vet Sci.* 2014;34:889–96.
- Marx C, Silveira MD, Beyer Nardi M. Adipose-Derived Stem Cells in Veterinary Medicine: Characterization and Therapeutic Applications. *Stem Cells Dev.* 2015;24:803-13.
- McClellan PD. A proposal for standardization in sonographic imaging I: metacarpus and metatarsus. *Equine Vet Sci.* 1986;6:327–28.
- Miles S, Price GM, Swift L, Shepstone L, Leinster SJ. Statistics teaching in medical school: opinions of practising doctors. *BMC Med Educ.* 2010;10:75.
- Munsterman AS, Kottwitz JJ, Reid Hanson R. Meta-Analysis of the Effects of Adhesion Barriers on Adhesion Formation in the Horse. *Vet Surg.* 2016;45:587-95.
- Murphy EA: *The Logic of Medicine.* Baltimore: Johns Hopkins University Press, 1976.
- Murray R, Dyson SJ, Tranquille C, Adams V. Association of type of sport and performance level with anatomical site of orthopaedic injury and injury diagnosis. *Equine vet J Suppl.* 2006;36:411-16.
- Muttini A, Valbonetti L, Abate M, Colosimo A, Curimo V, Mauro A, Berardinelli P, Russo V, Cocciolone D, Marchisio M, Mattioli M, Tosi U, Podaliri Vulpiani M, Barboni B. Ovine amniotic epithelial cells: in vitro characterization and transplantation into equine superficial digital flexor tendon spontaneous defects. *Res Vet Sci.* 2013;94:158–69.
- Muttini A, Russo V, Rossi E, Mattioli M, Barboni B, Tosi U, Maffulli N, Valbonetti L, Abate M. Pilot experimental study on amniotic epithelial mesenchymal cell transplantation in natural occurring tendinopathy in horses. Ultrasonographic and histological comparison. *Muscles Ligaments Tendons J.* 2015;5:5-11.
- Nixon AJ, Dahlgren LA, Haupt JL, Yeager AE, Ward DL. Effect of adipose-derived nucleated cell fractions on tendon repair in horses with collagenase-induced tendinitis. *Am J Vet Res.* 2008;69:928– 37.
- Odgaard-Jensen J, Vist GE, Timmer A, Kunz R, Akl EA, Schünemann H, Briel M, Nordmann AJ, Pregno S, Oxman AD. Randomisation to protect against selection bias in healthcare trials. *Cochrane Database Syst Rev.* 2011; 4 (doi:10.1002/14651858.MR000012.pub3).
- O’Meara B, Bladon B, Parkin TDH, Fraser B, Lischer CJ. An investigation of the relationship between race performance and superficial digital flexor tendonitis in the Thoroughbred racehorse. *Equine Vet J.* 2010;42: 322–26.
- Paccini S, Spinabella S, Trombi L, Fazzi R, Galimberti S, Dini F, Carlucci F, Petrini M. Suspension of bone marrow-derived undifferentiated mesenchymal stromal cells for repair of superficial digital flexor tendon in race horses. *Tissue Eng.* 2007;13:2949-55.
- Patterson-Kane JC, Wilson AM, Firth EC, Parry DAD, Goodship AE. Exercise-related alterations in crimp morphology in the central core regions of superficial digital flexor tendons from young Thoroughbreds: a controlled study. *Equine Vet J.* 1998;30:61–64.
- Pearson K. Report on Certain Enteric Fever Inoculation Statistics. *Br Med J.* 1904;5:1243-46.
- Personett L, McAllister E, Mansmann R. Proximal suspensory desmitis. *Mod Vet Pract.* 1983;64:541-45.
- Peters JL, Sutton AJ, Jones DR, Rushton L, Abrams KR. A Systematic Review of Systematic Reviews and Meta-Analyses of Animal Experiments with Guidelines for Reporting. *J Environ Sci Health B:Pesticides, Food Contaminants, and Agricultural Wastes.* 2006;41(7):1245-58.

- Petrisor BA, Bhandari M. Principles of teaching evidence-based medicine. *Injury*. 2006;37:335-9.
- Plevin S, Mc Lellan J. The effect of insertional suspensory branch desmitis on racing performance in juvenile Thoroughbred racehorses. *Equine Vet J*. 2014;46:451-7.
- Pool RR, Meagher DM. Pathologic findings and pathogenesis of racetrack injuries. *Vet Clin North Am Eq Pract*. 1990;6:1-30.
- Preuss M. Das Metaanalyse Modell nach DerSimonian & Laird mit exakten Gewichten [Dissertation]. Lübeck: Univ. Lübeck; 2013.
- Rantanen NW, Jorgensen JS, Genovese RL. Ultrasonographic Evaluation of the Equine Limb: Technique. In: Ross MW, Dyson SJ, Hrsg. *Diagnosis and Management of Lameness in the Horse*. 2. Aufl. St. Louis: Elsevier; 2011. p. 182-205.
- Reed SA, Leahy ER. Growth and Development Symposium: Stem cell therapy in equine tendon injury. *J Anim Sci*. 2013;91:59-65.
- Reef VB. Superficial digital flexor tendon healing: ultrasonographic evaluation of therapies. *Vet Clin North Am*. 2001;17:159-78.
- Renzi S, Ricco S, Dotti S, Sesso L, Grolli S, Cornalli M, Carlin S, Patruno M, Cinotti S, Ferrari M. Autologous bone marrow mesenchymal stromal cells for regeneration of injured equine ligaments and tendons: A clinical report. *Res Vet Sci*. 2013;95:272-77.
- Ressing M, Blettner M, Klug SJ. Systematische Übersichtsarbeiten und Metaanalysen. *Dtsch Ärztebl Int*. 2009;106:456-63.
- Ricco S, Renzi S, Del Bue M, Conti V, Merli E, Ramoni R, Lucarelli E, Gnudi G, Ferrari M, Grolli S. Allogeneic adipose tissue-derived mesenchymal stem cells in combination with platelet rich plasma are safe and effective in the therapy of superficial digital flexor tendonitis in the horse. *Int J Immunopathol Pharmacol (Suppl. 1)*. 2013;26:61-8.
- Rich FR. Single-Center Study of 83 Horses with Suspensory Injuries Treated with Adipose-Derived Stem and Regenerative Cells. *Stem Cell Discovery*. 2014;4:44-53.
- Rich T, Patterson-Kane JC. Science-in-brief: What is needed to prevent tendon injury in equine athletes? A conversation between researchers and industry stakeholders. *Equine Vet J*. 2014;46:393-98.
- Richardson LE, Dudhia J, Clegg PD, Smith R. Stem cells in veterinary medicine-attempts at regenerating equine tendon after injury. *Trends Biotechnol*. 2007;25:409e16.
- Rijkenhuizen AB, Gietemann IV, van der Reijd TC. Desmitis des Fesselträgers und Desmopathie des Fesselträgerursprungs an der Vorder- und Hintergliedmaße des Pferdes – eine retrospektive Studie über die Erfolge unterschiedlicher Behandlungen in der Praxis. *Pferdeheilkd*. 2007;23:357-65.
- Romagnoli N, Rinnovati R, Ricciardi G, Lambertini C, Spinella G, Spadari A. Clinical Evaluation of Intraleisional Injection of Platelet-Rich Plasma for the Treatment of Proximal-Suspensory Ligament Desmitis in Horses. *J Equine Vet Sci*. 2015;35:141-46.
- Rosenbrock A, Jacobi R, Jaugstetter H, Brehm W. Autologous bone marrow transplantation to stimulate suspensory ligament regeneration in 24 horses – A clinical case control study. 12th ESVOT Congress. 2004 Sep 10-12; Munich, Germany. p. 264-65.
- Russell JW, Russell TM, Vasey JR, Hall MS. Autologous bone marrow aspirate for treatment of superficial digital flexor tendonitis in 105 racehorses. *Vet Rec*. 2016;179:69.



- Ryan JM, Barry FP, Murphy JM, Mahon BP. Mesenchymal stem cells avoid allogeneic rejection. *J Inflamm (Lond)*. 2005;Jul 26:2-8.
- Sargeant JM, Kelton DF, O'Connor AM. Study designs and systematic reviews of interventions: building evidence across study designs. *Zoonoses Public Health*; 2014, Jun; 61 Suppl 1:10-7 (doi: 10.1111/zph.12127).
- Sawdon H, Yovich JV, Booth T. Superficial digital flexor tendonitis in racehorses: Long term follow up of conservatively managed horses. *Aust Equine Vet*. 1996;14:21-5.
- Scala M, Lenarduzzi S, Spagnolo F, Trapasso M, Ottonello C, Muraglia A, Barla A, Scillario M, Strada P. Regenerative medicine for the treatment of Teno-desmic injuries of the equine. A series of 150 horses treated with platelet-derived growth factors. *In vivo*. 2014;28:1119-23.
- Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB. Evidence-based medicine: what it is and what it isn't. *Br med J*. 1996;312:71-2.
- Schmidt PL. Evidence-based veterinary medicine: evolution, revolution, or repackaging of veterinary practice? *Vet Clin North Am Small Anim Pract* 2007;37 (3):409-417 (doi:10.1016/j.cvsm.2007.01.001).
- Schmidt H. Die Behandlung akuter und chronischer Sehnenerkrankungen beim Pferd mit hochmolekularer Hyaluronsäure [Dissertation med. vet.]. Hannover: Tierärztliche Hochschule; 1989.
- Schnabel LV. Use of Biologic Therapies for the Treatment of Tendon and Ligament Injuries. Proceedings of the NAVC Conference; Orlando; 2016 Jan 16-20, Florida, USA. 2016; Volume 29: 194-96.
- Schulz KF, Grimes DA. Sample size calculation in randomized trials: mandatory and mystical. *Lancet* 2005;365:1348-53.
- Schulze T, Budras K-D. Zur klinisch-funktionellen Anatomie des M. interosseus medius der Hintergliedmaße im Hinblick auf die Insertionsdesmopathie des Pferdes – Kernspin-, computertomographische- und morphologische Untersuchungen. *Pferdeheilkd*. 2008;24:343-50.
- Schwörer S, Becker F, Feller C, Baig AH, Köber U, Henze H, Kraus JM, Xin B, Lechel A, Lipka DB, Varghese CS, Schmidt M, Rohs R, Aebersold R, Medina KL, Kestler HA, Neri F, von Maltzahn J, Tümpel S, Rudolph KL. Epigenetic stress responses induce muscle stem cell aging by Hoxa9 developmental signals. *Nature* 2016;540:428-32.
- Serra Braganca FM, Rhodin M, van Weeren PR. On the brink of daily clinical application of objective gait analysis: What evidence do we have so far from studies using an induced lameness model? *Vet J*. 2018;234:11-23.
- Silver IA, Brown PN, Goodship AE, Lanyon LE, McCullagh KG, Perry GC, Williams IF. A clinical and experimental study of tendon injury, healing and treatment in the horse. *Equine Vet J Suppl* 1. 1983;1:1-43.
- Simoneit C, Heuwieser W, Arlt S. Die evidenzbasierte Veterinärmedizin im Praxisalltag. *Tierarztl Prax Ausg G Grosstiere Nutztiere*. 2012;40:186-92.
- Smith ML, Glass GV. Metaanalysis of psychotherapy outcome studies. *Am Psychol*. 1977;32:752-60.
- Smith RK. Pathophysiology of tendon injury. In: Ross MW, Dyson SJ, Hrsg. *Diagnosis and Management of Lameness in the Horse*. 1. Aufl. Philadelphia: Saunders; 2003 b. p. 616-28.
- Smith RK. Mesenchymal stem cells therapy for equine tendinopathy. *Disabil Rehabil*. 2008 a;30:1752-8.
- Smith RK. Stem cell therapy for tendon and ligament injuries in the horse – technique and outcome. *Pferdeheilkd*. 2008 b;24:513-18.

Smith RK. Rehabilitation of equine tendon injuries. Proceedings of the 25th Annual Scientific Meeting European College of Veterinary Surgeons; Centro Cultural de Belem; 2016 Jul 7-9, Lisbon, Portugal. Zürich: ECVS; 2016. p. 104-6.

Smith RK, Korda M, Blunn GW, Goodship AE. Isolation and implantation of autologous equine mesenchymal stem cells from bone marrow into the superficial digital flexor tendon as a potential novel treatment. *Equine Vet J*. 2003 a;35:99-102.

Smith RK, McIlwraith CW, Schweitzer R, Kadler K, Cook J, Caterson B, Dakin S, Heinegård D, Screen H, Stover S, Crevier-Denoix N, Clegg P, Collins M, Little C, Frisbie D, Kjaer M, van Weeren R, Werpy N, Denoix J-M, Carr A, Goldberg A, Bramlage L, Smith M, Nixon A. Advances in the understanding of tendinopathies: A report on the Second Havemeyer Workshop on equine tendon disease. *Equine Vet J*. 2014;46:4-9.

Smith RK, Schramme M. Tendon injury in the horse: Current theories and therapies. In *Practice*. 2003;25: 529-39.

Smith RK, Werling NJ, Dakin SG, Alam R, Goodship AE, Dudhia J. Beneficial effects of autologous bone marrow-derived mesenchymal stem cells in naturally occurring tendinopathy. *PLoS One*. 2013;8 (doi: 10.1371/journal.pone.0075697).

Smith RK, Young N, Dudhia J, Kasashima Y, Clegg P, Goodship A. Effectiveness of bone-marrow-derived mesenchymal progenitor cells for naturally occurring tendinopathy in the horse. *Regen Med (Suppl. 2)*. 2009;4:25.

Sodersten F, Hultenby K, Heinegård D, Johnston C, Ekman S. Immunolocalization of collagens (I and III) and cartilage oligomeric matrix protein in the normal and injured equine superficial digital flexor tendon. *Connect Tissue Res*. 2013;54:62-9.

Sousa NR de, Luna SPL, Pizzigatti D, Martins MTA, Possebon FS, Aguiar ACS. Relation between type and local of orthopedic injuries with physical activity in horses. *Ciencia Rural* 2017;47:1-7 (doi.org/10.1590/0103-8478cr20151218).

Sullivan SL, Whitem T, Morley PS, Hinchcliff KW. A systematic review and meta-analysis of the efficacy of furosemide for exercise-induced pulmonary haemorrhage in Thoroughbred and Standardbred racehorses. *Equine Vet J*. 2015;47:341-9.

Takahashi T, Kasashima Y, Ueno Y. Association between race history and risk of superficial digital flexor tendon injury in Thoroughbred racehorses. *J Am Vet Med Assoc*. 2004;225:90-3.

Tassan S, Cremonesi F, Sanzogni E, Perego R, Perrini C, Lange-Consiglio A. Clinical application of amniotic derived mesenchymal cells conditioned medium in tendon and ligament injuries of sport horses. *Ippologia*. 2014;25:3-14.

Tetta C, Lange-Consiglio A, Bruno S, Tetta E, Gatti E, Dobrev M, Cremonesi F, Camussi G. The role of microvesicles derived from mesenchymal stem cells in tissue regeneration; a dream for tendon repair? *Muscle Ligaments Tendons J*. 2012;2:212-21.

Textor JA. Autologous biologic treatment for equine musculoskeletal injuries: Platelet-rich plasma and IL-1 receptor antagonist protein. *Vet Clin North Am Equine Pract*. 2011;27:275-98.

Textor JA, Willits NH, Tablin F. Synovial fluid growth factor and cytokine concentrations after intra-articular injection of a platelet-rich product in horses. *Vet J*. 2013;198:217-23.

Thorpe CT, Clegg PD, Birch HL. A review of tendon injury: why is the equine superficial digital flexor tendon most at risk? *Equine Vet J*. 2010;42:174-80.

Timmer A, Rücker G. Systematische Übersichtsarbeiten zu Fragen der Therapie und Prävention. Teil 3 – Wie werden die Ergebnisse zusammengefasst und dargestellt? *AMT Arzneimitteltherapie*, 2008;26:299-303.

- Tipton TE, Ray CS, Hand DR. Superficial digital flexor tendonitis in cutting horses: 19 cases (2007–2011). *J Am Vet Med Assoc.* 2013;243:1162-65.
- Trump M. A retrospective study of the prevalence of injuries to the suspensory ligament, digital flexor tendons and associated structures in a non-racehorse referral-hospital population [Dissertation med. vet.]. Zürich: Uni Zürich; 2014.
- Vandenbergh A, Broeckx SY, Beerts C, Seys B, Zimmerman M, Verweire I, Suls M, Spaas HJ. Tenogenically induced allogeneic mesenchymal stem cells for the treatment of proximal suspensory ligament desmitis in a horse. *Front Vet Sci.* 2015;2:49.
- Van Loon VJ, Scheffer CJ, Genn HJ, Hoogendoorn AC, Greve JW. Clinical follow-up of horses treated with allogeneic equine mesenchymal stem cells derived from umbilical cord blood for different tendon and ligament disorders. *Vet Q.* 2014;34:92-7.
- Van Schie HT, Bakker EM, Jonker AM, van Weeren PR. Ultrasonographic tissue characterization of equine superficial digital flexor tendons by means of gray level statistics. *Am JV et Res.* 2000;61:210-9.
- Van Weeren PR, Pfau T, Rhodin M, Roepstorff L, Serra Braganca FM, Weishaupt M. Do we have to redefine lameness in the era of quantitative gait analysis? *Equine Vet J.* 2017;49:567-69.
- Verwilghen D, Caudron I, van Galen G, Gabriel A, Grulke S, Serteyn D. Equine tendonitis: a review on current therapies. Part 2. *Ann Méd Vét.* 2010;154:1-15.
- Wagner C. Metaanalyse zur Therapie der bovinen Dermatitis digitalis. [Dissertation med. vet.]. Gießen: Justus-Liebig-Universität Gießen; 2014.
- Wang JH. Mechanobiology of tendon, *J Biomech.* 2006;39:1563-82.
- Waselau M, Sutter W, Genovese RL, Bertone A. Intralesional Injection of Platelet-Rich Plasma Followed by Controlled Exercise for Treatment of Midbody Suspensory Ligament Desmitis in Standardbred Racehorses. *J Am Vet Med Assoc.* 2008;232:1515-20.
- Williams IF, Heaton A, McCullagh KG. Cell morphology and collagen types in equine tendon scar. *Res Vet Sci.* 1980;28:302-10.
- Williams RB, Harkins LS, Hammond CJ, Wood LJN. Racehorse injuries, clinical problems and fatalities recorded on British racecourses from flat racing and National Hunt racing during 1996, 1997 and 1998. *Equine Vet J.* 2001;33:478-86.
- Yovich JV, Sawdon H, Booth T, Robertson ID. Correlation of ultrasonographic findings and long term outcome in racehorses with superficial digital flexor tendon injury. *Aust Equine Pract.* 1995;13:89-92.
- Ziegler A, Lange S, Bender R. Systematische Übersichten und Metaanalysen – Artikel Nr. 19 der Statistik-Serie in der DMW –Systematic reviews and Metaanalyses. *Dtsch Med Wochenschr* 2007;132:48–52 (doi: 10.1055/s-2007-959042).
- Zuffova K, Krisova S, Zert Z. Platelet rich plasma treatment of superficial digital flexor tendon lesions in racing Thoroughbreds. *Vet Med-Czech.* 2013;58:230-39.
- Zwolanek D, Satué M, Proell V, Godoy JR, Odörfer KI, Flicker M, Hoffmann SC, Rüllicke T, Erben RG. Tracking mesenchymal stem cell contributions to regeneration in an immunocompetent cartilage regeneration model. *JCI Insight.* 2017;2 (doi: 10.1172/jci.insight.87322).

## DANKSAGUNG

Herrn Prof. Dr. Walter Brehm danke ich herzlich für die Überlassung dieses sehr interessanten Themas. Er hat mir vielfältige Anregungen und den notwendigen Gestaltungsspielraum gegeben es wissenschaftlich zu bearbeiten.

Dankbar bin ich Herrn Prof. Dr. Geburek für seine wissenschaftliche Arbeit, die mir den Einstieg in das Thema ermöglicht hat, und das vertiefende Fachgespräch bei der Fachfortbildung.

Frau Mag. med. vet. Doris Reinitzer von der Universitätsbibliothek VU Wien danke ich für ihre Hilfsbereitschaft, die sie mir über die Jahre mit großer fachlicher Kompetenz bei der wissenschaftlichen Literatursuche und Fragen zur Datenbanknutzung entgegenbrachte, sowie die interessanten Diskussionen über die evidenzbasierte Veterinärmedizin.

Herrn Markus Döbbeler, prakt. Tierarzt, danke ich für seinen fundierten medizinischen Rat und das entgegengebrachte Verständnis, dass ich neben der beruflichen Tätigkeit in seiner Tierarztpraxis auch wissenschaftlich eingebunden war.

Allen tierärztlichen Kolleginnen und Kollegen sei für die fachlichen Beiträge, die vertiefenden Gespräche und sachlichen Hinweise auf Kongressen und Fachtagungen gedankt.

Besonderer Dank gilt meinen Eltern, Maria Bärbel und Dr. Peter Doll, für Ihre Unterstützung in allen Lebenslagen und die uneingeschränkt positive Einstellung bis zum Abschluss dieser Arbeit.